The effect of photoaging on the accumulation of somatic mutations in cancer-free human skin

Barbara Hernando[1], Michelle Dietzen[2], Genis Parra[3], Marta Gil-Barrachina[1], Gerard Pitarch[4], Laura Mahiques[4], Francisca Valcuende-Cavero[5], Nicholas McGranahan[2,6,*], Conrado Martinez-Cadenas[1,*]

[1]Department of Medicine, Jaume I University of Castellon, Ave. Sos Baynat, 12071 Castellon, Spain
[2]Cancer Genome Evolution Research Group, University College London Cancer Institute, Huntley St. WC1E 6AG, London, United Kingdom
[3]Centre Nacional d'Anàlisi Genòmica-Centre de Regulació Genòmica (CNAG-CRG), Parc Científic de Barcelona, 08028 Barcelona, Spain.
[4]Department of Dermatology, Castellon University General Hospital. Ave. Benicassim, 12004, Castellon, Spain.
[5]Department of Dermatology, La Plana University Hospital. Carretera Villarreal-Burriana, 12540, Villarreal, Spain.
[6]Lead contact

*Correspondence: nicholas.mcgranahan.10@ucl.ac.uk (N.M.) and ccadenas@uji.es (C.M-C.)

SUMMARY

Sunlight is the principal aetiological factor associated with skin cancer development. However, genetic and phenotypic factors also contribute to skin cancer risk. In this study, we deeply sequenced 46 cancer genes in normal skin biopsies from 127 healthy donors. Our results reveal an exponential accumulation of UV-related somatic mutations with age, matching skin cancer incidence. The increase of mutational burden is in turn modified by an individual's skin phototype. Somatic mutations preferentially accumulated in cSCC driver genes and clonally expanded with age, with distinct mutational processes underpinning different age groups. Our findings reveal that aged, sun-exposed normal skin is an extended mosaic of multiple clones with driver mutations, poised for the acquisition of transforming events.

KEY WORDS

Somatic mutation; Aging; Skin Phototype; Next-generation sequencing; UV exposure; normal epidermis; carcinogenesis; mutational spectrum

SIGNIFICANCE

This study delineates how somatic mutations accumulate in sun-exposed normal skin from different age and skin phototype groups. Having a large cohort has allowed us to explore the effect of well-known intrinsic and extrinsic factors in the mutational burden of normal skin. Apart from age, an individual's skin phototype is key in the build-up of somatic mutations in healthy skin, playing a more significant role than the history or the body site pattern of sun exposure. This could be of vital value for the development of specific clinical strategies to manage patients with increased risk of skin cancer. Additionally, our results highlight the importance of discriminating clones with tumorigenic potential from benign clones with mutations that will not cause a functional impairment of the tissue.

INTRODUCTION

All cancers arise as a result of somatic alterations occurring in the genomic DNA sequence of 'normal' cells. Somatic genomic alterations accumulate spontaneously in cells throughout a person's life as a result of errors occurring during cell replication or after exposure to mutagenic agents, such as certain chemicals in tobacco smoke or UV radiation from sunlight (Martincorena and Campbell, 2015; Stratton et al., 2009).

The accumulation of clones of cells harbouring mutations across tissues may be expected to have functional consequences to the physiology of normal cells, contributing to ageing and promoting disease progress, as in the case of cancer. In this regard, recent sequencing studies have revealed that, in general, the somatic mutational burden and profile of physiological normal tissues seem to be relatively similar to those found in tumours from the same tissue (Brunner et al., 2019; Jaiswal et al., 2014; Lee-Six et al., 2019; Martincorena, 2019). These results suggest that the majority of somatic alterations may appear randomly, pre-dating tumours, and only a small fraction of all somatic mutations in a cancer genome are, therefore, relevant in carcinogenesis, disease classification and treatment.

In the case of skin, a recent study analysing cancer-free epidermal samples from four individuals showed that the frequency of driver mutations in physiologically normal skin cells is surprisingly high (Martincorena et al., 2015). Sun-exposed epidermal cells carried a multitude of genetic alterations, and about 25-30% of these normal skin cells had already acquired at least one driver mutation, indicating that cancer-causing mutations are under strong positive selection even in epidermis maintaining normal physiological functions. The mutational profile found in eyelid samples, a chronically sun-exposed area, was similar to cutaneous squamous cell carcinomas (SCCs), but distinct from the profile characteristically observed in cutaneous melanoma (Martincorena et al., 2015). This may be due to the low number of epidermal melanocytes (cells from which melanoma develops) in relation to keratinocytes (Hoath and Leahy, 2003), but also due to the fact that melanoma appears more frequently in sporadically, rather than chronically, sun-exposed areas of the body – those that are usually covered by clothing. This apparently paradoxical fact is attributable to injuries caused by an intermittent pattern of intense and acute sun exposure associated with recreational activities (Gandini et al., 2005).

Skin cancer incidence worldwide reveals a clear relationship between pigmentation traits and sunlight damage, with individuals with fair skin and inability to tan showing greater cancer susceptibility. Cutaneous sensitivity to sunlight exposure (ability to tan *versus* tendency to burn) is defined by certain genetically determined pigmentation traits (Scherer and Kumar, 2010). Thus, individuals carrying genetic variants associated with increased sun sensitivity – and with low skin phototype – should have higher somatic mutational rates, as they have reduced melanin levels, insufficient to protect the genome of epidermal cells from the mutagenic action of UV light (Brenner and Hearing, 2008; Rijken et al., 2004).

While previous studies have confirmed that seemingly normal cells harbour mutations, the key factors that determine which individuals are more prone to acquire and therefore accumulate somatic mutations remains unclear. This is probably due to the fact that most studies have explored somatic alterations in a limited number of individuals. With the aim of increasing our understanding of the accumulation somatic mutations in the skin as a consequence of different patterns of sunlight exposure, we focused on sequencing skin samples from a large cohort of 127 cancer-free individuals (Figure 1). We examined somatic mutational burdens, mutational signatures, clonal selection, and frequency of driver mutations in normal epidermal samples from chronically-, intermittently- and non-photoexposed body sites. Furthermore, we explored if there is an added risk to accumulate somatic mutations according to intrinsic characteristics of individuals (i.e. age, and pigmentation-related genotype and phenotype) and to sun exposure habits.

RESULTS

Variation of mutational burden across samples

We performed an ultra-deep targeted sequencing of 46 cancer genes in normal epidermal samples obtained from 127 cancer-free volunteers, aged 10 to 92 years (mean of 58.07 years), undergoing an excision of benign cutaneous lesions. The average on-target coverage across samples was 953.82x (range 377.96-1657.37x). Skin samples were collected from different body areas, classified according to the pattern of sunlight exposure as (a) chronically-photoexposed (n=44), (b) intermittently-photoexposed (n=79), and (c) non-photoexposed (n=4). Each sample was collected from one donor due to ethical reasons. The study cohort is described in Table S1.

A total of 5,301 mutations were identified in our dataset, with an average of 41.74 mutations per sample (range 2-169) (Figure 2A), and at an average rate of 130.56 mutations per megabase (range 6.25-528.13). The mutational burden variation across samples significantly correlated with age (Spearman's ρ=0.61, *P*-value = 2.89×10^{-14}; Figure 2B) and pattern of body site photoexposure (Kruskal-Wallis test, *P*-value = 1.12×10^{-4}; Figure 2C). Consistent with previous studies of normal tissue, most mutations were likely present only in a small fraction of cells, evidenced by the fact that the majority of mutations exhibited a variant allele frequency (VAF) lower than 5% (Figure S1D).

The size of our cohort offered us a unique opportunity to directly quantify the key factors associated with mutational burden in sun-exposed skin samples. Indeed, several phenotypic and behavioural risk factors were collected from the participants, including sex, age, Fitzpatrick's skin phototype, history of sunlight exposure, body site pattern of sun exposure, signs of sun damage in the skin area biopsied and *MC1R* genotype.

To empirically assess the relative importance of each potential risk factor in contributing to the mutational burden of skin lesions, we used a log-linear multivariate model. Strikingly, the total variance of mutational burden explained by all variables combined was high (adjusted-R^2 = 49.88%), with age explaining the largest proportion of the total variance (55.16%; Figure 3A). Surprisingly, this analysis revealed that an individual's skin phototype is the second strongest predictor, explaining 17.92% of the mutational burden variance across samples. Our results suggest a significant decrease ($\beta < 0$) in the number of somatic mutations accumulated in skin samples from individuals with high skin phototypes (skin types that normally tan after sunlight exposure), as compared with individuals with low skin phototypes (skin types that normally burn after sunlight exposure) (Table S2). Therefore, the ability to protect the skin against UV radiation plays a key role in both photoaging and likely skin cancer. This is consistent with the fact that people with low skin phototypes (I/II) have a tendency to develop sunburns and frequently lack the ability to acquire a tan following exposure to sunlight. This lack of tanning capacity confers greater susceptibility to develop skin cancers, due in part to the inability to protect against UV-related DNA damage (Scherer and Kumar, 2010).

Contrary to previous studies (Robles-Espinoza et al., 2016), no association was found between the somatic mutational load and the genotype in the melanocortin 1 receptor (*MC1R*), a key pigmentation-related gene determining the ability to respond to UV exposure (García-Borrón et al., 2014). Moreover, a lack of significant association between mutational burden and the pattern of body site photoexposure (chronic *versus* intermittent) was observed after including all risk factors in the model, explaining only 7.82% of the variance of the total number of mutations accumulated across samples. Our data suggest a comparable effect of intermittent exposure to sunlight, perhaps while on recreational activities, and continuous exposure, through

spending a large amount of daylight time outdoors, with regards to the accumulation of somatic mutations in normal epidermis.

Finally, given that age and phototype were the two most significant contributors to explain mutational burden, we quantitatively assessed age-related mutational burden according to skin phototype. For each skin phototype, we determined the mutation rate increased per year of life using a log-linear model (Figure 3B). Nonparametric bootstraps (1000 runs) were conducted to estimate the 95% confidence intervals (CI95) of the age effect. Somatic accumulation rates per year of life in the selected genes (0.32 Mb) increased across skin phototypes, ranging from 1.07 mutations per year for type I (6.72%, CI95: 5.86-7.68%) to 1.04 for skin phototype IV (4.17%, CI95: 3.28-4.76%).

Taken together, these results suggest a profound influence of age on the accumulation of somatic mutations in normal skin. This age-associated rise of mutational burden is in turn modified by an individual's skin phototype, indicating the inability of UV-sensitive individuals to protect against UV-related DNA damage.

Aging and the rise of UV-associated mutations

To explore the mutational processes underpinning the accumulation of mutations in sun-exposed samples, we considered the specific substitution types. In the cohort as a whole, we observed a predominance of C>T and CC>TT mutations at dipyrimidine sites (TpC or CpC context), likely reflecting repair of 6,4-photoproducts and the production of cyclobutane pyrimidine dimers (CPDs) in response to UV-induced DNA damage (Ravanat et al., 2001) (Figure 4A). These C>T substitutions were preferentially accumulated on the non-transcribed to the transcribed strand (Poisson test, *P*-value = 4.21×10^{-4}), consistent with the activity of transcription-coupled nucleotide excision repair (Figure 4B). In addition, we also observed an enrichment of T>C/A>G mutations at CTT sites, potentially caused by indirect DNA damage after UV radiation – greater incorporation of G, rather than A, opposite thymidine and cytidine photodimers by translesion polymerases (Figure 4A). This is in line with the fact that these mutations tended to accumulate in the transcribed rather than in the non-transcribed strand (Poisson test, *P*-value = 5.43×10^{-5}) (Figure 4B).

To quantify the presence of specific mutational signatures, which may reflect underlying mutational processes, we applied deconstructSigs to the combined set of mutations within the cohort. Non-photoexposed skin samples were discarded for signature analyses because of the limited number of samples and the small number of mutations. The specific mutational spectrum observed in our samples mainly reassembles the COSMIC base substitution signatures related to UV radiation (mutational signatures SBS7a-d). In fact, 44.32% of mutations could be attributed to the SBS7b signature, which may reflect the existence of mutational processes related to dipyrimidine photoproducts, with signatures SBS7c and SBS17a accounting for 13.92% and 13.76% of mutations, respectively (Figure S2B).

Remarkably, the relative contribution of UV-related mutational pathogenesis varied markedly with age (Figure 4C). Mutational signatures directly related to UV contributed 68.14% of all mutations detected after the age of 63 (the cohort median age), and only 46.59% of the mutational burden in individuals younger than 63 (Figure S2B). This observation is in accordance with the fact that non-melanoma skin tumours (BCC and cSCC) typically occur at advanced age and are related to cumulative sun exposure (Leiter et al., 2014). In fact, we observed that the increase of UV-mutations in normal skin with age follows a similar exponential trend than skin cancer incidence in Spain (data downloaded from the Global Cancer Observatory, http://gco.iarc.fr) (Figure 4D). The main difference between the elderly and younger individuals in terms of mutational spectra, apart from the fraction of UV-related mutations, was related to the fraction of T>C/A>G mutations at CTT contexts, being surprisingly high in younger individuals. This mutational pattern closely matched the SBS17a

signature, a COSMIC signature with unknown aetiology that has been shown to contribute to cutaneous melanoma (Alexandrov et al., 2019). To gain insight into the context dependency of this T>C/A>G substitutions, we explored the local sequence contexts (from -5 to +5 positions) and observed a specific pattern of contextual preference (CTTTT) in normal skin samples biopsied from younger individuals (Figure S3A). Additionally, the degree of transcriptional strand bias detected for T>C/A>G substitutions was substantially higher in younger individuals (Figure S3B). These differences in the mutational spectra are mainly caused by the exponential increase of C>T substitutions with age, since the absolute number of T>C/A>G mutations are similar among age groups. These observations suggest that initially the acquisition of T>C/A>G substitutions in normal skin is relatively high. However, after the age of 60, the mutagenic process related to sunlight is the major contributor to the accumulation of somatic mutations in skin.

To confirm that the differences shown in mutational spectra with age are not due to unequal sun-exposure between subgroups (the proportion of chronically-photoexposed skin samples is higher in the elderly group than in the young group), we implemented deconstructSigs splitting samples of each age group according to the pattern of sunlight exposure of the skin tissue biopsied. These results confirmed that the sunlight exposure of the skin sample was not a confounding factor in the age-specific mutational pattern observed (Figure S3C).

Taken together, our results reveal that distinct mutational processes operate with age. After an initial accumulation at younger ages of the SBS17a signature, which seems to remain steady during life, UV-related mutational processes appear to cause more mutations as age progresses. This suggests a clear decline in the ability to repair UV induced mutagenic lesions later in life.

Positive selection of driver mutations in normal skin

We next considered whether protein-altering somatic mutations were subjected to positive selection. We evaluated the footprint of positive selection in two orthogonal ways, by quantifying the excess of non-synonymous mutations as well as by estimating clone size.

The majority of the normal sun-exposed skin samples harboured multiple protein-altering mutations, even though these epidermal samples were histologically benign (Figure S4A). Given the significant effect of age in mutational burden observed, together with the fact that clonal expansions of cancer-associated mutations are extremely common with age (Risques and Kennedy, 2018), we examined if these mutations accumulate randomly or if there is also an enrichment of non-synonymous mutations with age. As expected, the number of putative driver mutations accumulated per sample increased with age (Wilcoxon-Mann-Whitney test, *P*-value = 1.10×10^{-9}; Figure 5A), with the non-synonymous mutational burden in samples from elderly individuals even higher than the one observed in cSCC (Wilcoxon-Mann-Whitney test, *P*-value = 2.40×10^{-5}) and in cutaneous melanoma samples (Wilcoxon-Mann-Whitney test, *P*-value = 2.50×10^{-9}; Figure 5A). In addition, these non-synonymous mutations tended to accumulate in specific cancer-associated genes (Figure 5B). We found that the catalogue of recurrently mutated genes in normal skin was almost identical to that of cSCC, with *TP53, NOTCH1*, *NOTCH2* and *FAT1* being recurrently mutated in both normal skin and cSCC samples (Figure S4B).

In order to quantify the extent of positive selection driving minor clonal expansion in normal skin samples in both elderly and young donors, we considered the ratio of missense, nonsense and essential splicing mutations compared to synonymous mutations using the dNdScv package. Our results provided evidence of significant positive selection when considering mutations in all cancer genes as a whole (Figure 5C) and specifically in three out of 46 sequenced genes for normal skin samples collected from both young and elderly individuals (*TP53, NOTCH1,* and *FAT1*) (Figure 5D). Notably, all three genes were also shown to have a significant excess of non-synonymous mutations in normal sun-exposed epidermis (Martincorena et al., 2015), and

have been shown to be drivers of cutaneous squamous cell carcinoma (cSCC) (Inman et al., 2018; Pickering et al., 2014).

Additionally, a significant excess of truncating point mutations in *NOTCH2* and *CDKN2A*, two tumour suppressor genes recurrently mutated in skin cancers (Hayward et al., 2017; Inman et al., 2018; Pickering et al., 2014), were also found in samples from elderly individuals. In addition, we identified samples with canonical hotspot mutations with therapeutic relevance (according to the database of curated mutations) in several oncogenes such as *BRAF* (V600E), *HRAS* (G12D, G12V), *PIK3CA* (P471L, E542K, E726K, M1043I, H1047L), and *FGFR3* (R248C, S249C, Y373C, A391E). In total, forty-two donors (81% of them older than 63 years) in our cohort carried at least one of the 50 different relevant disease-causing mutations identified.

In order to confirm whether these protein-altering mutations are exerting a selective pressure for clonal expansion, we scrutinized the variant allele frequency (VAF) distribution of somatic mutations in each age group. Overall, variant allele frequencies of non-synonymous mutations significantly increased in epidermal samples from elderly donors compared to those collected from young donors (Wilcoxon-Mann-Whitney test, *P*-value = 8.30×10^{-8}; Figure 5E), suggesting tolerance and selection for larger clones with age.

To further quantify whether non-synonymous mutations had an appreciably impact on cell fitness, we compared the sizes of clones carrying non-synonymous mutations to those with synonymous mutations. The average VAF of non-synonymous mutations per gene was significantly higher than that of synonymous mutations in normal skin samples from both young (Wilcoxon-Mann-Whitney test, *P*-value = 9.96×10^{-6}) and elderly donors (Wilcoxon-Mann-Whitney test, *P*-value = 6.21×10^{-8}; Figure 5E). The highest average VAF was observed in cancer-associated genes under positive selection (*NOTCH1, FAT1* and *TP53*), suggesting that these selectively advantageous mutations may appear early and expand with age in sun-exposed normal skin. However, the relative low frequency of somatic mutations also pointed out that they are present in only a small subset of skin cells, therefore remaining in a minority of sun-exposed cells.

A high number of sun-exposed samples (79.67%) carried at least one non-synonymous mutation in a cancer gene under positive selection (*TP53*, *NOTCH1*, and *FAT1*). Contrary to our expectation, no differences in the distribution of VAF (normalized by the average VAF of all other detected mutations per sample) were found when comparing skin samples carrying only one mutation with those carrying multiple different mutations in one of the three positively selected genes (Figure S5A). Therefore, together with findings previously found, clone size seems to be closely related to age. Consistent with previous findings (Yizhak et al., 2019), samples harbouring mutations in *TP53*, *NOTCH1* and *FAT1*, as well as other canonical hotspot mutations, had a significant increase in the overall number of mutations. Although this increase seems to be largely influenced by the participant's age, mutation count was age-independently associated with carrying protein-altering mutations in these positively selected genes (Figure S5B).

Taken as a whole, our data suggest that protein-affecting mutations in 5 of the 46 sequenced genes (*TP53, NOCTH1, FAT1, NOTCH2* and *CDKN2A*) are subject to statistically significant positive selection. Furthermore, we find that skin from elderly individuals not only harbours more mutations, but a larger fraction of these reflects positive selection. We speculate that the majority of the cells carrying driver mutations may not have yet acquired the right combination of mutated genes for expanding and culminating in the development of malignancy. In this regard, although recurrently mutated genes were similar between normal and cSCC samples, the distribution of variant frequencies in normal skin was significantly lower than the one observed in skin cancer subtypes, especially for driver genes (Figure S5C).

DISCUSSION

Like other tissues, skin undergoes chronological aging that might lead to its functional decline, which is accelerated by chronic sun damage. This study aimed to explore the role of photoaging, as well as other well-known epidemiological risk factors, in the accumulation of somatic mutations in cancer-free human epidermis. Our experimental design was focused on analysing the mutational landscape in a large cohort of subjects with a wide range of ages and phenotypic characteristics.

A profound variability in terms of mutational burden and driver mutations was observed among individuals. This variability was largely explained by age; on average 3.25 mutations/Mb were acquired in the panel of 46 genes per year of life. In addition, normal skin samples collected from donors with low skin phototypes (individuals with fair skin who tend to burn rather than tan after being exposed to sunlight) tended to accumulate a higher number of somatic mutations over time. Fair-skinned individuals are more severely affected by photoaging (Fisher et al., 2002). Therefore, the increased risk of developing skin cancer, at least for sporadic cancers, associated with these phenotypes may be due to the presence of a higher reservoir of mutant cells waiting to acquire more cancer-driving mutations, evade clonal growth control and initiate malignant transformation. By contrast, sun exposure habits were not significantly correlated to mutational burden variability across samples. Note that our samples are collected from Spanish individuals living in a region with relatively high UV index and pleasant weather throughout the year. Perhaps a higher effect of sun exposure pattern by body site would be found in populations from more northerly latitudes, exposed to sunlight mainly during summer vacations. Furthermore, the retrospective and subjective nature of some behavioural risk factors, particularly those related to sun exposure habits, raises the potential for recall bias. Therefore, validation in independent and larger cohorts will be needed to further analyse the association between behavioural risk factors and mutation burden.

Analysing a large cohort with a wide age range has allowed us to investigate the timing of somatic mutation accumulation in normal skin. Nearly all T>C mutations are accumulated early in life, as the number of these mutations accumulated in normal skin from young and elderly donors is similar. In contrast, an exponential rise of UV-related mutations with increasing age was evident in normal epidermis. This exponential increase of UV-related mutations, together with the fact that these mutations are mainly accumulated in the coding strand, could be related to the decline of nucleotide excision repair (NER) capacity with increasing age (Gorbunova et al., 2007). NER is a highly evolutionarily conserved mechanism for repairing bulky DNA lesions resulting, among others, from sunlight exposure. The importance of NER activity in the prevention of skin cancer is denoted by the extreme sensitivity to sunlight and severe predisposition to UV-induced skin cancers of patients with the inherited disorder xeroderma pigmentosum, in which genes encoding for the different components of the NER cascade are mutated (Lehmann et al., 2018).

Remarkably, the age-dependent exponential increase of skin cancer incidence in Spain recapitulates the accumulation of UV-related mutations in normal skin. Apart from the decline of NER function, skin photoaging has also been linked to impaired skin homeostasis (Panich et al., 2016). It is thought that strategies for tissue maintenance have a noteworthy impact on cancer incidence, consistent with the dramatic increase of cancer incidence with tissue and stem cell fitness decline (Rozhok and DeGregori, 2016). Aged/damaged tissue microenvironment may therefore provide an opportunity for clones with a selective advantage to expand, and that may be the reason why mutational profile notably differs between normal skin samples from young and adult donors.

Our results revealed an enrichment of driver mutations in the majority of normal sun-exposed skin samples, especially in those collected from elderly individuals. In addition, there is a marked overrepresentation of protein-altering mutations in several cSCC driver genes,

especially in *NOTCH1*, *TP53* and *FAT1*, likely reflecting positive selection. However, because these genes have been shown to be frequently mutated in normal skin (Martincorena et al., 2015; Yizhak et al., 2019), it seems unlikely that these mutations alone confer a sufficient growth advantage to engender cancer development. Indeed, although clones carrying these selectively advantageous mutations have expanded in normal skin, it is notable that only 21 out of 127 donors (16.53%) carried more than one mutation with clinical relevance in their skin. These mutant clones also seem to be relatively small, suggesting even in these cases the somatic alterations are often in distinct clones.

Cellular senescence, an irreversible proliferative arrest triggered by exogenous and endogenous stresses, may be a plausible explanation for the limited expansion of clones carrying cancer-causing mutations in normal tissues. Oncogenic-induced senescence is considered a crucial protective mechanism against cell transformation. Several reports from animal models support the idea that cell senescence may occur in tissues after acquiring a mitogenic mutation, preventing carcinogenesis at as initial step (Braig et al., 2005; Collado et al., 2005). The biology of naevi is a clear example of cellular senescence following an initial activating oncogenic mutation, normally in *BRAF* or *NRAS* genes (Bennett, 2003; Bennett and Medrano, 2002). Melanocytic nevi are clonal proliferations of non-malignant melanocytic cells, which can remain non-growing for many years, but also can act as precursors of melanoma if cells overcome senescence. Interventions that favour oncogene-induced senescence may help restrict the growth of clones carrying cancer-causing mutations in normal tissues and thus tumour progression.

The question arising from our observations, together with those shown in different sequencing studies previously performed on healthy tissues (Brunner et al., 2019; Jaiswal et al., 2014; Lee-Six et al., 2019; Martincorena et al., 2015, 2018; Suda et al., 2018), is whether or not targeting these early mutations recurrently found in normal tissues will be relevant for preventing carcinogenesis. Further efforts should be done to delineate the succession of genetic alterations needed for malignant transformation of physiologically normal tissues to premalignant precursor lesions, and finally to tumours, with the aim of discriminating drivers of the disease from the non-pathogenic mutational landscape. Regarding skin cancers, studying the mechanism of clonal evolution in normal skin offers the possibility to develop strategies for early detection and treatment of putative tumorigenic clones, but also for prevention of skin cancer development, especially in high-risk individuals such as people with risky skin phototypes.

METHODS

Sample collection

A total of 127 cancer-free volunteers donated a normal skin sample obtained from the margin of skin excision biopsies undertaken to remove a cutaneous benign lesion (Table S1). Four samples were collected from non-photoexposed skin areas (gluteus and armpit), 79 samples were collected from skin areas with intermittent sun exposure (back, chest, legs and upper arms), and 44 samples were obtained from chronically sun-exposed skin areas (neck, face and hands). Samples were recruited at the Department of Dermatology of two hospitals from Castellon Province, Spain (Castellon University General Hospital and La Plana University Hospital). Only one sample was recruited per donor due to ethical reasons. All participants provided a written informed consent. The study was approved by the Ethics Committee of the Jaume I University of Castellon (Spain).

Immediately after resection, tissue samples were submerged in RNA*later* Tissue Collection Solution (Thermo Fisher Scientific, Walham, MA, USA) and stored at 4ºC overnight. Then, the epidermis was separated from the dermis by incubating the tissue sample in 3.8% ammonium

thiocyanate (Sigma-Aldrich, St Louis, MO, USA) in PBS (pH 7.4) at room temperature for 3 hours. Subsequently, the epidermis was immersed in RNA*later* solution and stored at -20ºC until sample processing.

Genomic DNA was isolated from fresh-frozen normal epidermal samples with the QIAamp DNA Mini Kit, (Qiagen, Hilden, Germany). DNA was stored at -20ºC until use.

Phenotypic data collection

Each participant completed a standardised questionnaire to collect information on sex, age, pigmentation traits (skin, hair and eye colour), skin sensitivity to sunlight (tanning ability *versus* tendency to burn), freckling degree, history of childhood sunburns, and sun exposure habits. Pigmentation- and sun sensitivity-related traits were used to group individuals according to Fitzpatrick's skin type classification. Detailed information related to signs of sun damage in the skin area biopsied (pigmented spots, blotches, and wrinkles) was also recorded. To avoid misclassification, each participant completed the questionnaire under the supervision of a professional.

Sequencing of *MC1R* coding region

The coding sequence of the *MC1R* gene was directly sequenced in all samples, as previously described (Martínez-Cadenas et al., 2013). Non-synonymous *MC1R* mutations were then defined as 'R', 'r' or 'p' (pseudoallele) alleles according to their impact on protein function, following criteria previously described (Hernando et al., 2018).

Ultra-deep targeting sequencing

A panel of 46 genes was chosen to perform ultra-deep targeted sequencing. These genes have been found to be often involved in skin cancer development (Hayward et al., 2017; Inman et al., 2018; Jayaraman et al., 2014) and/or have been shown to be frequently mutated in normal skin samples (Martincorena et al., 2015). A custom bait capture was designed using NimbleGen SeqCap EZ (Roche, Basel, Switzerland) in order to target the exonic regions of the selected genes. The total size of the targeted regions was 0.32 Mb.

Sequencing of paired-end 100bp reads was performed on an Illumina HiSeq 2000 machine. The average on-target coverage across samples was 953.82x, ranging from 377.96x to 1657.37x. The variation in coverage across genes and samples is displayed in Figure S6. Note that the mutation burden found per gene, as well as per sample, was not strongly influenced by differences in coverage (Figure S6C).

Paired-end reads were aligned to the reference human genome (GRCh37d5) using the BWA-MEM algorithm with default parameters (Li and Durbin, 2009). Alignment files (BAM format) containing only properly paired, uniquely mapping reads were processed using Picard tools version 1.110 (http://broadinstitute.github.io/picard/) to add read groups and remove PCR duplicates. Local realignments and base-quality recalibrations were conducted using GATK (v.3.2.2) (McKenna et al., 2010).

Variant calling

Processed BAM files were analysed to identify single-nucleotide variants (SNVs) and small insertions and deletions (indels). Somatic mutations are normally called by detecting mismatches present in a tissue sample that are absent in a matched control sample (normal tissue or blood from the same patient). Due to the absence of matched normal sample from each individual, processed BAM files were used to perform somatic variant calling by applying Mutect2 in tumour-only mode (version 4.0.8.1). Following Broad Institute recommendations for

variant calling, putative artefacts were removed with FilterMutectCalls and FilterByOrientationBias. We provided FilterMutectCalls the set of human variants from gnomAD (https://gnomad.broadinstitute.org). Functional annotations were added to the resulting list of variants using SnpEff (Cingolani et al., 2012a), with the gene annotation based on Ensembl data release 75. Variants were then annotated using SnpSift (Cingolani et al., 2012b), with population frequencies, conservation scores and deleteriousness predictions obtained from dbNSFP (Liu et al., 2013). Each variant was also annotated using gnomAD, COSMIC, ExAC, and ClinVar.

Then, a number of post-processing filters were applied. Firstly, we focused on identifying and removing germline variants. The variant caller Platypus was used to identify germline variants, which were filtered out from the list of somatic mutations (Rimmer et al., 2014). The tool was run using the human variant set from dbSNP as the reference, instead of using a matched normal sample. Mutations detected in each sample were additionally called against the aggregate variants from a panel of normal samples of 200 Spanish individuals sequenced in the facilities of the CNAG-CRG (Barcelona, Spain) in order to remove common single-nucleotide polymorphisms (SNPs) and frequent technical artefacts. Indeed, variants were filtered out if they appeared in any of the ExAC, 1000 Genomes Project and dbSNP databases. As our filtering strategy seems to be quite rigorous, we decided to not remove those mutations that are included in the catalogue of somatic mutations found in human cancers (according to COSMIC and DoCM databases) for downstream analyses. These 75 putative driver mutations had a low VAF (mean = 0.021, max = 0.093) and prevalence (mean = 1.42%, max = 2.74%) in our cohort. To reduce false positive calls, variants were also filtered out based on their allele frequency. Our study was designed to detect mutations present in a small fraction of the skin cells of the biopsy. Variant allele frequencies for somatic mutations in normal samples are more likely to have values below 50%, as shown previously in normal skin samples from eyelids (Martincorena et al., 2015). Therefore, we filtered out a variant when the 95% confidence intervals (CIs) of its VAF (determined by the binomial distribution taking into account the depth of coverage) reached values greater than 50%. To increase the sensitivity of our analyses, we also opted to remove all variants present with VAF values two standard deviations away from the mean per sample. As we were working with a small cohort of unrelated patients, mutations with a prevalence higher than 25% in our cohort were further excluded. Next, we also excluded variants that have not been found to have a clinical relevance in human cancers (according to DoCM database) with a prevalence two standard deviations away from the mean of our cohort. That is because spontaneously-arising neutral mutations are extremely unlikely to affect samples collected from different patients. Finally, sites with very low coverage (n<50 reads) were also excluded to avoid testing sites with limited power to detect variants.
To check if the majority of mutations removed were germline variants or technical artefacts, we studied the context-specific mutation spectra for each set of mutations removed at each filtering step (Figure S1A). Note that the majority of substitutions removed were not related to UV damage (C>T mutations at dipyrimidine sites). Indeed, having a global dN/dS ratio << 1 may denote that the pre-filtering dataset of variants is contaminated with germline SNPs (Figure S1B).

Prediction of mutational burden

A log-linear model was applied to correlate the number of mutations detected per sample with the sun exposure pattern of the skin sample and the individual's age, including different phenotypic traits as covariates. The covariates included in the model were sex (female *vs.* male), skin phototypes (I *vs.* II, III or IV), sun damage in the tissue (absence *vs.* presence), history of sunlight exposure (frequently *vs*. occasionally), and *MC1R* genotype (wild-type *vs.* r carriers or R carriers). The R package 'relaimpo' was used to assess the relative importance of the different variables included in the model. Non-photoexposed skin samples were excluded in this analysis due to the reduced sample size of this group (n=4). Nonparametric bootstraps (1000 runs) were

conducted to estimate the 95% confidence intervals (CI95) of the age effect in each skin phototype subgroup.

In order to double-check that the majority of mutations were real somatic variants, we applied the log-linear model using the pre-filtering dataset of variants (Figure S1C). Note that the total variance of mutational burden explained by the model was very low (adjusted-R^2 = 6.08%), and the major predictors of mutational burden were completely different than those obtained when the filtering mutation dataset was used.

Analysis of local mutational context and extraction of mutational signatures

Mutational spectrum and signatures analyses were performed by using the deconstructSigs R package (Rosenthal et al., 2016). Due to the limited number of mutations found in some samples (less than 50 mutations), we decided to group the samples by (a) age of individuals, and (b) pattern of sunlight exposure of skin tissue biopsied. Firstly, the proportion of each distinct single base substitution, as well as of each dinucleotide mutation, occurring within a given trinucleotide context was determined per each sample and per group. Hierarchical clustering of samples based on trinucleotide context of mutations was performed by applying the Ward's criterion. Samples were mainly divided into the different clusters by age and body site exposure, confirming that the vast majority of mutations included in our final list are real somatic mutations (Figure S2A).

Then, we evaluated the transcriptional strand bias for the mutations that are located within exons. A Poisson test was applied to assess whether the mutations occurred more often in the transcribed or untranscribed strand, or vice versa. Exon definitions for human reference genome were retrieved from BiomaRt by loading a TxDb annotation package from Bioconductor (Durinck et al., 2005).

The limited number of variants hampers the discovery of new mutational signatures. Therefore, we ran deconstructSigs including only the mutational signatures related to aging and/or previously observed in the different skin cancers subtypes that contribute at least 6% of all of the observed mutations across the 127 samples (SBS2, SBS6, SBS7a, SBS7b, SBS7d, and SBS17a). Figure S2B shows the weights assigned to all of these mutational signatures for the combined set of mutations within cohort (Total) and per age group. Due to the limited number of samples and the low number of mutations, non-photoexposed skin samples were excluded.

Prevalence of non-synonymous mutations and selection analyses

Selection across the normal skin samples was quantified by using the dNdScv R package (Martincorena et al., 2017), which adapts the traditional implementation of dNdS ratio by using trinucleotide context-dependent substitution models to avoid common mutation biases affecting dN/dS. Selection tests were performed on different subsets of mutations by grouping samples per age. Briefly, global and gene-level dNdS ratios were quantified for missense and truncating (nonsense and essential splicing) mutations, as well as for indels, and then were used to compare the selection intensities between skin samples biopsied from elderly and young individuals. Again, the fact that our results reveal an excess of non-synonymous mutations (dNdS > 1), especially in genes frequently involved in skin cancer development, suggests that our list of mutations is not contaminated by germline variants or technical artefacts (Figure 5C).

The database of curated mutations (DoCM, docm.genome.wustl.edu) was used to identify canonical hotspot mutations with characterized functional or clinical evidence in cancer.

Identification of frequently mutated genes in skin cancer

The dNdS package was also applied to detect recurrently mutated genes in publicly available skin cancer mutation calls of 40 cutaneous squamous cell carcinomas (Inman et al., 2018) and 140 cutaneous melanomas (Hayward et al., 2017). Then, the mutational landscape in skin cancer samples was compared to the observed in our dataset of normal skin samples. The analysis was restricted to the 46 genes selected for this study.

ACKNOWLEDGMENTS

We are extremely grateful to all the volunteers for giving their consent to take part in this study, as well as to all the medical specialists for supervising phenotype collection of all samples. This work is supported in part by the Jaume I University of Castellon (UJI-A2016-13). B.H. is funded by the Jaume I University of Castellon under a Postdoctoral Research contract (POSDOC-A/2018/07), and received additional funding for conducting a research exchange at the UCL Cancer Institute (E-2019-34). N.M is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (Grant Number 211179/Z/18/Z), and also receives funding from CRUK, Rosetrees, and the NIHR BRC at University College London Hospitals, and the CRUK University College London Experimental Cancer Medicine Centre.

AUTHOR CONTRIBUTIONS

B.H. treated skin tissues, performed DNA extraction, prepared samples for sequencing, conducted bioinformatics analysis, and wrote the manuscript. M.D. helped with mutational signatures analysis. G.P. helped with variant calling and filtering. M.G-B. helped with lab work. G.P, L.M. and F.V-C collected tissue samples and phenotypic data of donors. N.M. provided expertise in interpreting the results, supervised bioinformatics analyses, and wrote the manuscript. C.M-C conceived the project, supervised the study, and wrote the manuscript. All co-authors provided critical feedback and contributed to manuscript preparation.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. Schematic overview of the experimental design.

Figure 2. Somatic mutation burden in normal epidermis. (A) Total number and type of somatic mutations detected across the 46 genes sequenced in each sample. Clinical and demographic characteristics are presented below. **(B)** Correlation between donor age and the number of somatic mutations accumulated in the donor's sample. The fitted line, confidence interval, correlation coefficient (ρ), and *P*-value were obtained by performing a Spearman's correlation test. **(C)** Differences in the number of somatic mutations of normal epidermal samples according to the sunlight exposure pattern of body site from which the sample was resected. A Kruskal-Wallis test is used for testing differences among groups, and a Wilcoxon-Mann-Whitney test is used as a post-hoc test for performing pairwise comparisons.

Figure 3. Linear modelling of the accumulation of somatic mutations in normal epidermis. (A) Relative importance of predictors included in the log-linear regression model. The total variance explained by the model (adjusted-R^2 = 49.88%) is decomposed in order to know the individual contribution (effect size) of each predictor. Asterisk denotes significant predictors. **(B)** A log-linear regression is used for analysing the age effect in the accumulation of somatic

mutations for each skin phototype. Solid lines represent the bootstrapped mean of the slope, and shaded areas its bootstrapped 95% confidence intervals (CI95).

Figure 4. Spectrum of somatic mutations in normal epidermis. (A) Bar plot showing the fraction of single (top) and doublet (bottom) base substitutions found in each of the possible 96 trinucleotide context (strand independent). **(B)** Relative number of each substitution type present on the transcribed (dark shading) and untranscribed strand (light shading). Asterisks indicate significant transcriptional strand asymmetries (Poisson test). **(C)** Mutational spectra in samples from young (left) and elderly donors (right). Heatmaps show the fraction of each trinucleotide change in each sample (middle). Bar plots represent the contribution mean of each 96-mutation type per age group (top). Clinical and demographic characteristics are presented next to each sample (right). **(D)** Age-dependent increase of both UV-mutation accumulation and skin cancer incidence. Data of skin cancer incidence in Spain was downloaded from the Global Cancer Observatory (http://gco.iarc.fr).

Figure 5. Positive selection and expansion of driver mutations in normal skin with age. (A) Number of mutations detected per sample. Dots are coloured according to donor's age. A Wilcoxon-Mann-Whitney test is used for testing differences among groups. **(B)** Number of mutations detected per gene normalized by sample size of the respective group. A Wilcoxon-Mann-Whitney test is used for testing differences among groups. **(C)** Global dN/dS values estimated by taking together all 46 genes in normal skin biopsied from both young and elderly donors. Error bars represent 95% confidence intervals. **(D)** dN/dS ratios for each of the 46 target genes. Genes under significant positive selection in both age groups are coloured in red, while genes positively selected only in the elderly group are coloured in blue (overall q-value < 0.05). Genes are sorted from higher (bottom) to lower (top) significant value in the elderly group. **(E)** Distribution of VAFs of somatic non-synonymous and synonymous mutations per gene. Red dots denote positively selected genes in both young and elderly groups (*FAT1*, *NOTCH1*, and *TP53*), and blue dots indicate genes under positive selection only in the elderly group (*NOTCH2* and *CDKN2A*). Dots representing the other genes sequenced are coloured in grey. A Wilcoxon-Mann-Whitney test is used for testing differences among mutation types in each age group.

SUPPLEMENTAL INFORMATION

Supplemental information included six figures and two tables

Figure S1. Evaluation of variant calling and filtering, Related to Figures 2 to 5. (A) Spectra of mutation sets removed after applying a specific filtering step. All mutational spectra are very different from the typical UV-related mutational spectrum, indicating that the filtered variants are unlikely to be real somatic mutations. **(B)** Global dN/dS ratios estimated before and after mutation filtering called with Mutect2 tumour-only mode. The global dN/dS << 1 denotes contamination of germline variants and/or technical artefacts in the non-filtered dataset of somatic mutations. This problem seems to be solved after applying the different filtering steps (dN/dS > 1). Error bars denote 95% confidence interval. **(C)** Results of applying a log-linear regression model in the non-filtered mutation dataset for predicting the number of mutations per sample. The low variance explained by the model (adjusted-R^2 = 6.08%) denotes that the non-filtered list of mutations includes a large number of likely false positive calls. **(D)** Histogram of somatic mutations identified by variant allele frequency (VAF). Most somatic mutations remain in a subclonal state with low VAFs (VAF << 5%).

Figure S2. Mutational spectra in normal skin, Related to Figure 4. (A) Heatmap showing the fraction of each 96-mutation type per sample. Clinical and demographic characteristics are presented above each sample. **(B)** Percentage of substitutions attributed to each one of the six mutational signatures for all mutations from all 127 samples together (Total), as well as for all mutations included in each age subgroup.

Figure S3. Age-related mutational spectra in normal skin, Related to Figure 4. (A) Local mutational context of T>C substitutions in samples biopsied from young and elderly donors. **(B)** Relative number of each substitution type present on the transcribed (dark shading) and untranscribed strand (light shading) in samples biopsied from young and elderly donors. Asterisks indicate significant transcriptional strand asymmetries (Poisson test). **(C)** 96-barplot depicting the number of mutations observed at each trinucleotide context taking together all samples biopsied from young and elderly individuals (Total), as well as splitting samples of each age group by the body site pattern of sun exposure (Chronically- and Intermittently-photoexposed).

Figure S4. Occurrences of somatic mutations in the 46 cancer genes across samples, Related to Figure 5. (A) Heatmap showing the distribution of recurrent non-synonymous mutations per coding kilobase of sequence for each one of the 46 genes targeted across all normal skin samples. Clinical and demographic characteristics are presented above each sample. The *KMT2B* gene is not included in this plot since no non-synonymous mutation was found across samples. **(B)** Percentage of normal skin samples, as well as in cSCC and melanoma tumours, carrying at least one non-synonymous mutation in each gene.

Figure S5. Clonal expansion of clones with oncogenic mutations, Related to Figure 5. (A) Distribution of VAFs for *NOTCH1*, *TP53* and *FAT1* in normal skin samples carrying only one or multiple mutations in the respective gene. For samples with multiple mutations in one of these three genes, the mean VAF for all mutations per gene is plotted. Each dot represents a sample and is coloured according to the donor's age. In each panel, a Wilcoxon-Mann-Whitney test is used for testing differences among groups. **(B)** Number of non-synonymous mutations per sample in normal skin samples non-carriers or carriers of one or multiple non-synonymous mutations in *NOTCH1*, *TP53* and *FAT1*, as well as in normal skin without or with canonical hotspot mutations. Each dot represents a sample and is coloured according to the donor's age. For avoiding the confounding effects of age, samples were stratified according to donor's age for statistical analyses. In panels comparing more than two groups, a Kruskall-Wallis (KW) test is used for testing differences among groups. In panels comparing two groups, a Wilcoxon-Mann-Whitney (WMW) test is used for testing differences among groups. **(C)** Heatmap showing the mean VAF of all non-synonymous mutations found per gene across all normal, cSCC and melanoma samples.

Figure S6. Coverage and mutational burden across genes and samples, Related to Figures 2 to 5. (A) Plot showing the number of mutations per gene across all samples (bar plot, top) and the mean coverage per gene and sample (box plot, bottom). Genes in the x-axis sorted by mean coverage across samples. Blue line indicates the mean coverage across all samples. **(B)** Plot showing the number of mutations per sample (bar plot, top) and the mean coverage per sample (bar plot, bottom). Samples in the x-axis sorted by mean coverage across all sequenced regions. Blue line indicates the mean coverage across all samples. **(C)** Scatter plots showing the coverage and number of mutations per gene (left) and per sample (right). These plots show that coverage did not significantly influence the number of mutations found across genes and/or across samples.

Table S1. Demographic and clinical data of all Spanish donors, Related to Figures 1 to 5.

Table S2. Log-linear modelling of the accumulation of somatic mutations in normal skin, Related to Figure 3

REFERENCES

Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2019). The Repertoire of Mutational Signatures in Human Cancer. BioRxiv 322859.

Bennett, D.C. (2003). Human melanocyte senescence and melanoma susceptibility genes. Oncogene *22*, 3063–3069.

Bennett, D.C., and Medrano, E.E. (2002). Molecular regulation of melanocyte senescence. Pigment Cell Res. *15*, 242–250.

Braig, M., Lee, S., Loddenkemper, C., Rudolph, C., Peters, A.H.F.M., Schlegelberger, B., Stein, H., Dörken, B., Jenuwein, T., and Schmitt, C.A. (2005). Oncogene-induced senescence as an initial barrier in lymphoma development. Nature *436*, 660–665.

Brenner, M., and Hearing, V.J. (2008). The Protective Role of Melanin Against UV Damage in Human Skin. Photochem. Photobiol. *84*, 539–549.

Brunner, S.F., Roberts, N.D., Wylie, L.A., Moore, L., Aitken, S.J., Davies, S.E., Sanders, M.A., Ellis, P., Alder, C., Hooks, Y., et al. (2019). Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. Nature *574*, 538–542.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012a). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.

Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., and Lu, X. (2012b). Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front. Genet. *3*, 35.

Collado, M., Gil, J., Efeyan, A., Guerra, C., Schuhmacher, A.J., Barradas, M., Benguría, A., Zaballos, A., Flores, J.M., Barbacid, M., et al. (2005). Tumour biology: senescence in premalignant tumours. Nature *436*, 642.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinforma. Oxf. Engl. *21*, 3439–3440.

Fisher, G.J., Kang, S., Varani, J., Bata-Csorgo, Z., Wan, Y., Datta, S., and Voorhees, J.J. (2002). Mechanisms of Photoaging and Chronological Skin Aging. Arch. Dermatol. *138*, 1462–1470.

Gandini, S., Sera, F., Cattaruzza, M.S., Pasquini, P., Picconi, O., Boyle, P., and Melchi, C.F. (2005). Meta-analysis of risk factors for cutaneous melanoma: II. Sun exposure. Eur. J. Cancer Oxf. Engl. 1990 *41*, 45–60.

García-Borrón, J.C., Abdel-Malek, Z., and Jiménez-Cervantes, C. (2014). MC1R, the cAMP pathway, and the response to solar UV: extending the horizon beyond pigmentation. Pigment Cell Melanoma Res. *27*, 699–720.

Gorbunova, V., Seluanov, A., Mao, Z., and Hine, C. (2007). Changes in DNA repair during aging. Nucleic Acids Res. *35*, 7466–7474.

Hayward, N.K., Wilmott, J.S., Waddell, N., Johansson, P.A., Field, M.A., Nones, K., Patch, A.-M., Kakavand, H., Alexandrov, L.B., Burke, H., et al. (2017). Whole-genome landscapes of major melanoma subtypes. Nature *545*, 175–180.

Hernando, B., Ibañez, M.V., Deserio-Cuesta, J.A., Soria-Navarro, R., Vilar-Sastre, I., and Martinez-Cadenas, C. (2018). Genetic determinants of freckle occurrence in the Spanish population: Towards ephelides prediction from human DNA samples. Forensic Sci. Int. Genet. *33*, 38–47.

Hoath, S.B., and Leahy, D.G. (2003). The organization of human epidermis: functional epidermal units and phi proportionality. J. Invest. Dermatol. *121*, 1440–1446.

Inman, G.J., Wang, J., Nagano, A., Alexandrov, L.B., Purdie, K.J., Taylor, R.G., Sherwood, V., Thomson, J., Hogan, S., Spender, L.C., et al. (2018). The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature. Nat. Commun. *9*, 3667.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., et al. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med. *371*, 2488–2498.

Jayaraman, S.S., Rayhan, D.J., Hazany, S., and Kolodney, M.S. (2014). Mutational landscape of basal cell carcinomas by whole-exome sequencing. J. Invest. Dermatol. *134*, 213–220.

Lee-Six, H., Olafsson, S., Ellis, P., Osborne, R.J., Sanders, M.A., Moore, L., Georgakopoulos, N., Torrente, F., Noorani, A., Goddard, M., et al. (2019). The landscape of somatic mutation in normal colorectal epithelial cells. Nature *574*, 532–537.

Lehmann, J., Seebode, C., Martens, M.C., and Emmert, S. (2018). Xeroderma Pigmentosum - Facts and Perspectives. Anticancer Res. *38*, 1159–1164.

Leiter, U., Eigentler, T., and Garbe, C. (2014). Epidemiology of skin cancer. Adv. Exp. Med. Biol. *810*, 120–140.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

Liu, X., Jian, X., and Boerwinkle, E. (2013). dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. Hum. Mutat. *34*, E2393-2402.

Martincorena, I. (2019). Somatic mutation and clonal expansions in human tissues. Genome Med. *11*, 35.

Martincorena, I., and Campbell, P.J. (2015). Somatic mutation in cancer and normal cells. Science *349*, 1483–1489.

Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B., Tubio, J.M., et al. (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science *348*, 880–886.

Martincorena, I., Raine, K.M., Gerstung, M., Dawson, K.J., Haase, K., Van Loo, P., Davies, H., Stratton, M.R., and Campbell, P.J. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. Cell *171*, 1029-1041.e21.

Martincorena, I., Fowler, J.C., Wabik, A., Lawson, A.R.J., Abascal, F., Hall, M.W.J., Cagan, A., Murai, K., Mahbubani, K., Stratton, M.R., et al. (2018). Somatic mutant clones colonize the human esophagus with age. Science *362*, 911–917.

Martínez-Cadenas, C., López, S., Ribas, G., Flores, C., García, O., Sevilla, A., Smith-Zubiaga, I., Ibarrola-Villaba, M., Pino-Yanes, M. del M., Gardeazabal, J., et al. (2013). Simultaneous purifying selection on the ancestral MC1R allele and positive selection on the melanoma-risk allele V60L in south Europeans. Mol. Biol. Evol. *30*, 2654–2665.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

Panich, U., Sittithumcharee, G., Rathviboon, N., and Jirawatnotai, S. (2016). Ultraviolet Radiation-Induced Skin Aging: The Role of DNA Damage and Oxidative Stress in Epidermal Stem Cell Damage Mediated Skin Aging. Stem Cells Int. *2016*, 7370642.

Pickering, C.R., Zhou, J.H., Lee, J.J., Drummond, J.A., Peng, S.A., Saade, R.E., Tsai, K.Y., Curry, J.L., Tetzlaff, M.T., Lai, S.Y., et al. (2014). Mutational landscape of aggressive cutaneous squamous cell carcinoma. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *20*, 6582–6592.

Ravanat, J.L., Douki, T., and Cadet, J. (2001). Direct and indirect effects of UV radiation on DNA and its components. J. Photochem. Photobiol. B *63*, 88–102.

Rijken, F., Bruijnzeel, P.L.B., van Weelden, H., and Kiekens, R.C.M. (2004). Responses of black and white skin to solar-simulating radiation: differences in DNA photodamage, infiltrating neutrophils, proteolytic enzymes induced, keratinocyte activation, and IL-10 expression. J. Invest. Dermatol. *122*, 1448–1455.

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S.R.F., WGS500 Consortium, Wilkie, A.O.M., McVean, G., and Lunter, G. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat. Genet. *46*, 912–918.

Risques, R.A., and Kennedy, S.R. (2018). Aging and the rise of somatic cancer-associated mutations in normal tissues. PLoS Genet. *14*, e1007108.

Robles-Espinoza, C.D., Roberts, N.D., Chen, S., Leacy, F.P., Alexandrov, L.B., Pornputtapong, N., Halaban, R., Krauthammer, M., Cui, R., Timothy Bishop, D., et al. (2016). Germline MC1R status influences somatic mutation burden in melanoma. Nat. Commun. *7*, 12064.

Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. *17*, 31.

Rozhok, A.I., and DeGregori, J. (2016). The evolution of lifespan and age-dependent cancer risk. Trends Cancer *2*, 552–560.

Scherer, D., and Kumar, R. (2010). Genetics of pigmentation in skin cancer--a review. Mutat. Res. *705*, 141–153.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719–724.

Yizhak, K., Aguet, F., Kim, J., Hess, J.M., Kübler, K., Grimsby, J., Frazer, R., Zhang, H., Haradhvala, N.J., Rosebrock, D., et al. (2019). RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. Science *364*.

Figure 1

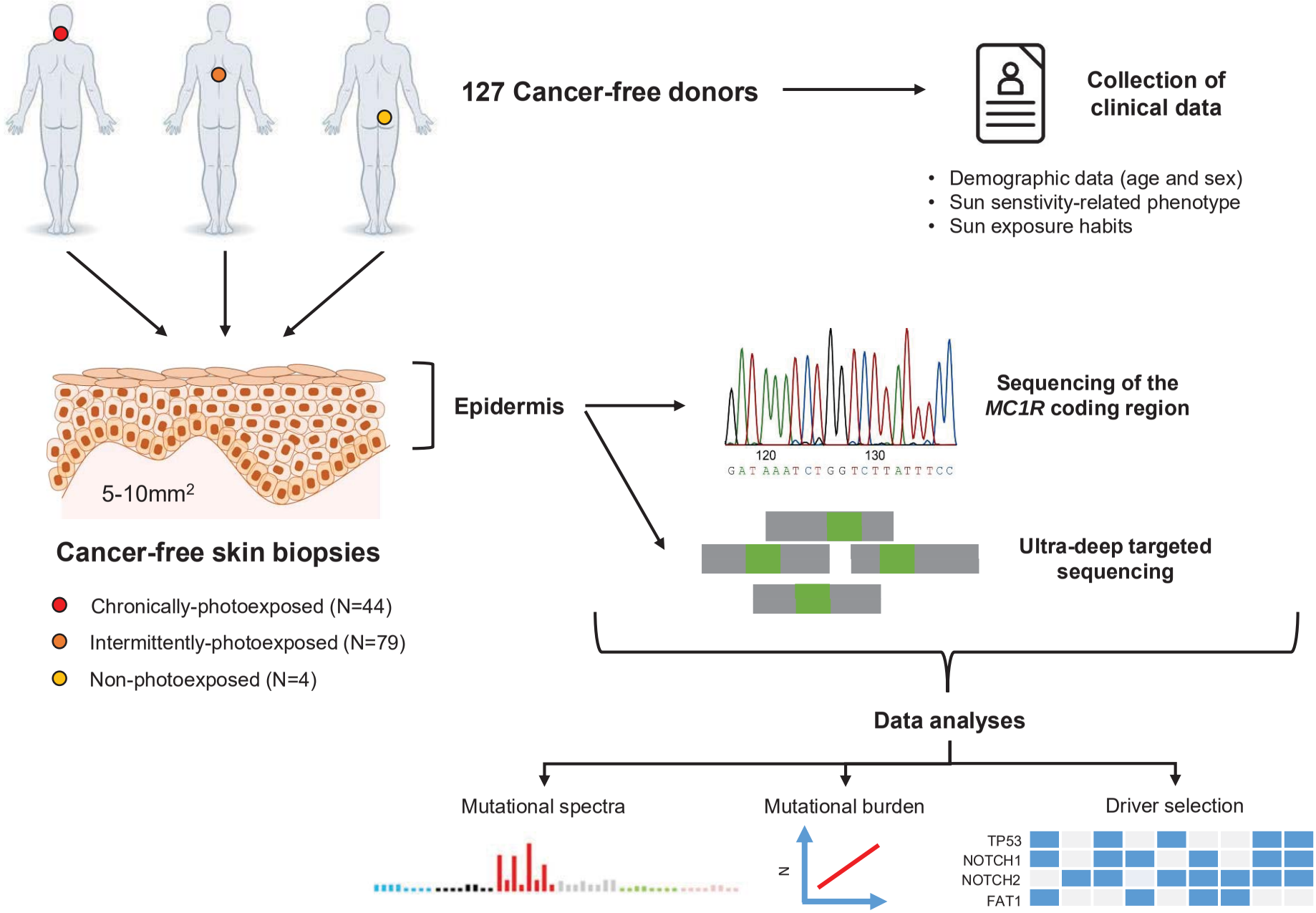

127 Cancer-free donors
Collection of clinical data
• Demographic data (age and sex)
• Sun senstivity-related phenotype
• Sun exposure habits
Epidermis
5-10mm2
Cancer-free skin biopsies
Chronically-photoexposed (N=44)
Intermittently-photoexposed (N=79)
Non-photoexposed (N=4)
120
130
GAT AAAT CT GG TCT TAT T TCC
Sequencing of the MC1R coding region
Ultra-deep targeted sequencing
Data analyses
Mutational spectra
Mutational burden
N
Driver selection
TP53
NOTCH1
NOTCH2
FAT1

Figure 2

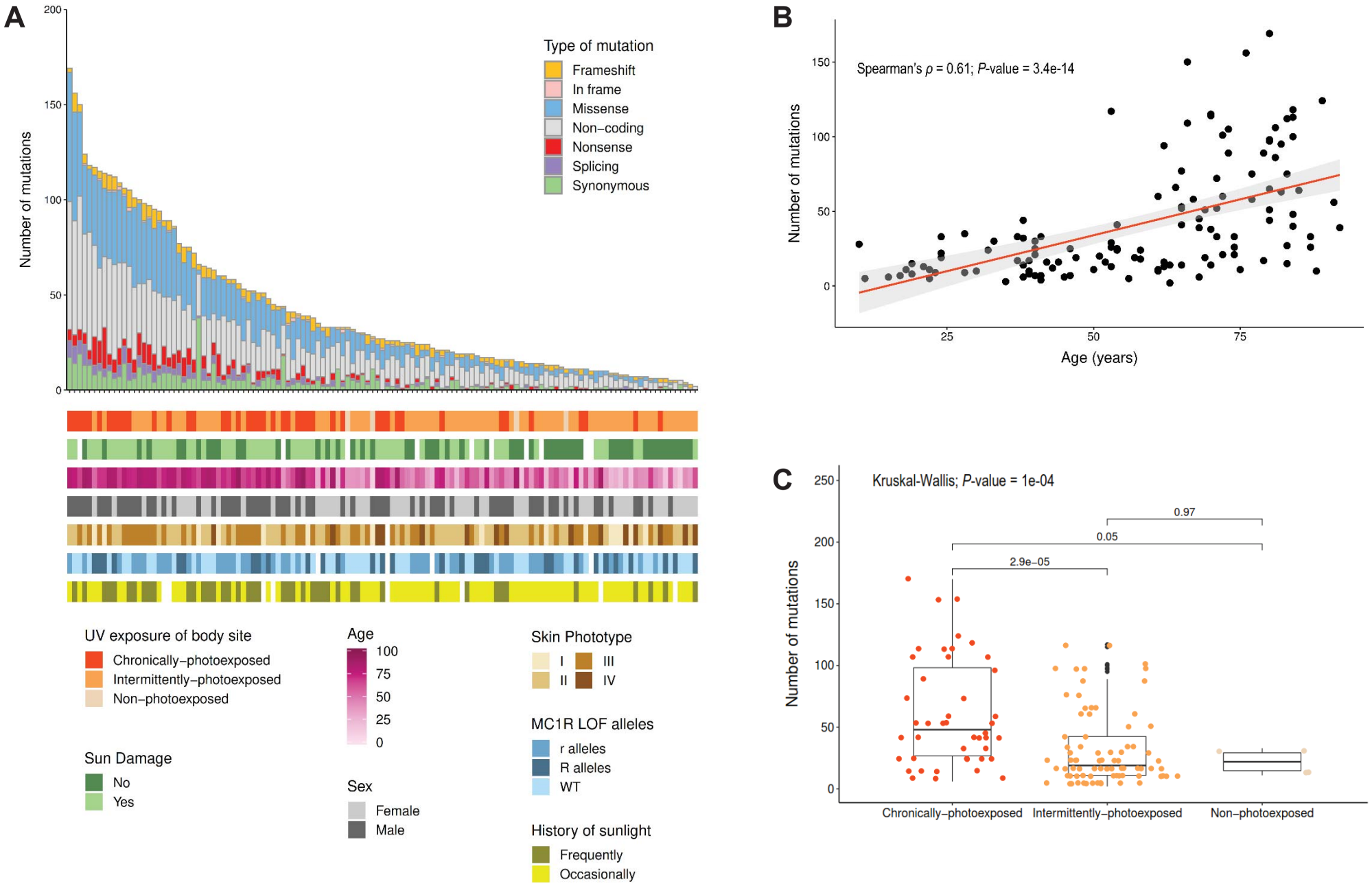
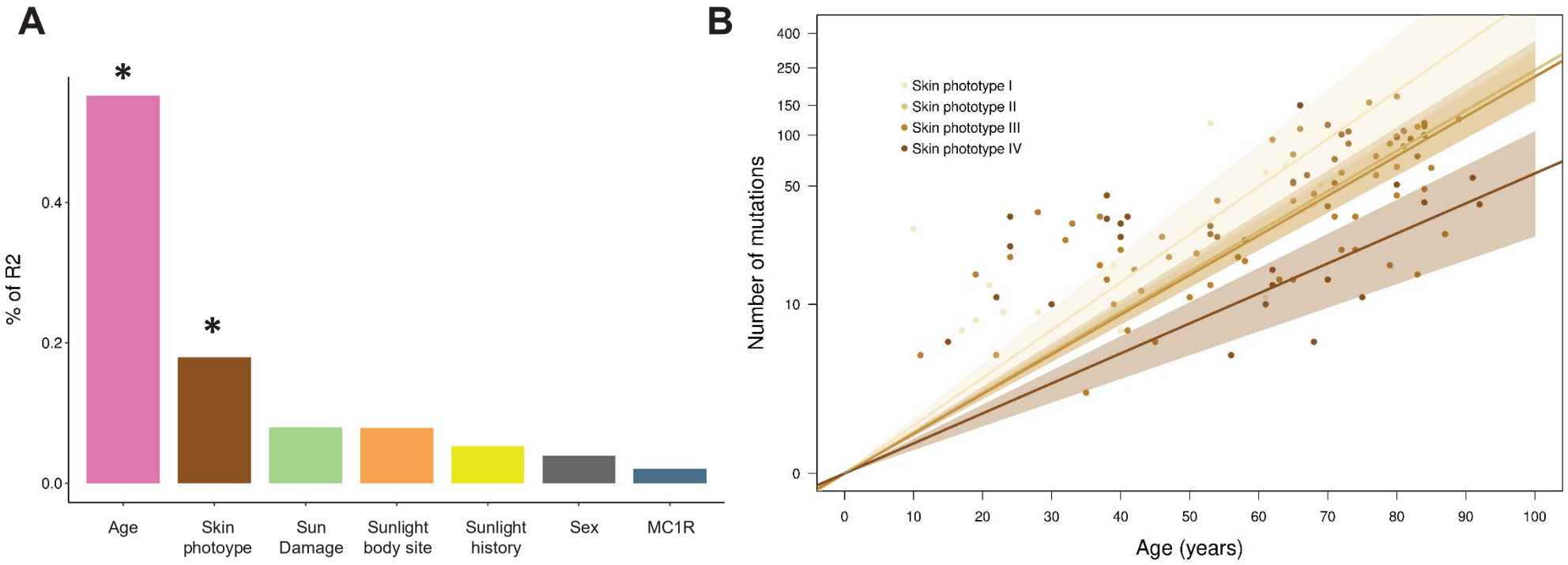
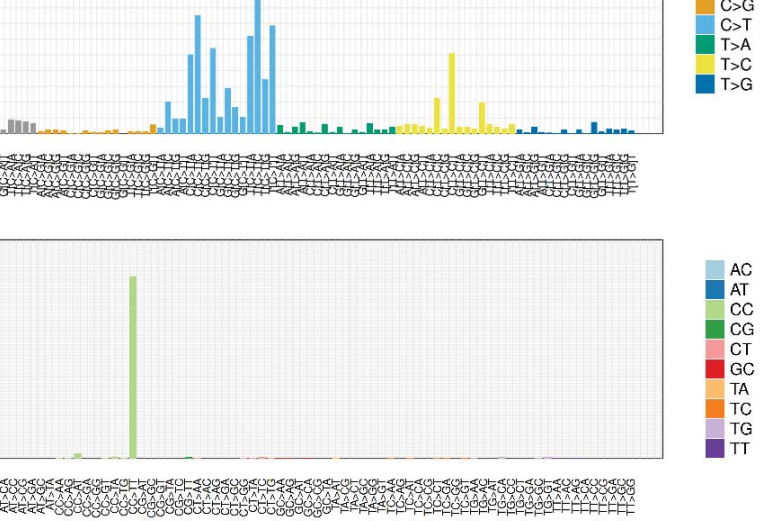
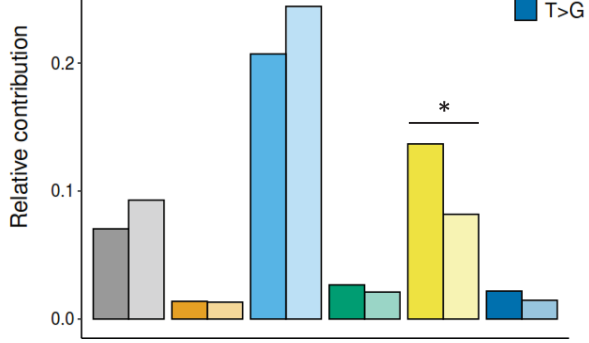
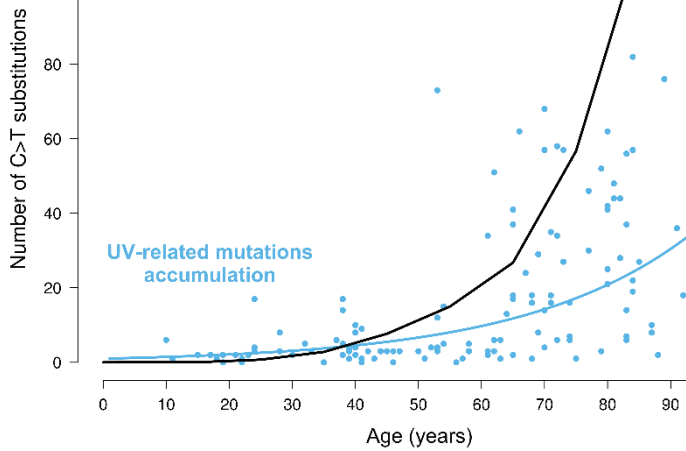

A
Number of mutations
200
150
100
50
0
Type of mutation
Frameshift
In frame
Missense
Non−coding
Nonsense
Splicing
Synonymous
UV exposure of body site
Chronically−photoexposed
Intermittently−photoexposed
Non−photoexposed
Sun Damage
No
Yes
Age
100
75
50
25
0
Sex
Female
Male
Skin Phototype
I
II
III
IV
MC1R LOF alleles
r alleles
R alleles
WT
History of sunlight
Frequently
Occasionally
B
Spearman's ρ = 0.61; P-value = 3.4e-14
Number of mutations
150
100
50
0
25
50
75
Age (years)
C
Kruskal-Wallis; P-value = 1e-04
0.97
0.05
2.9e−05
Number of mutations
250
200
150
100
50
0
Chronically−photoexposed
Intermittently−photoexposed
Non−photoexposed

Figure 3

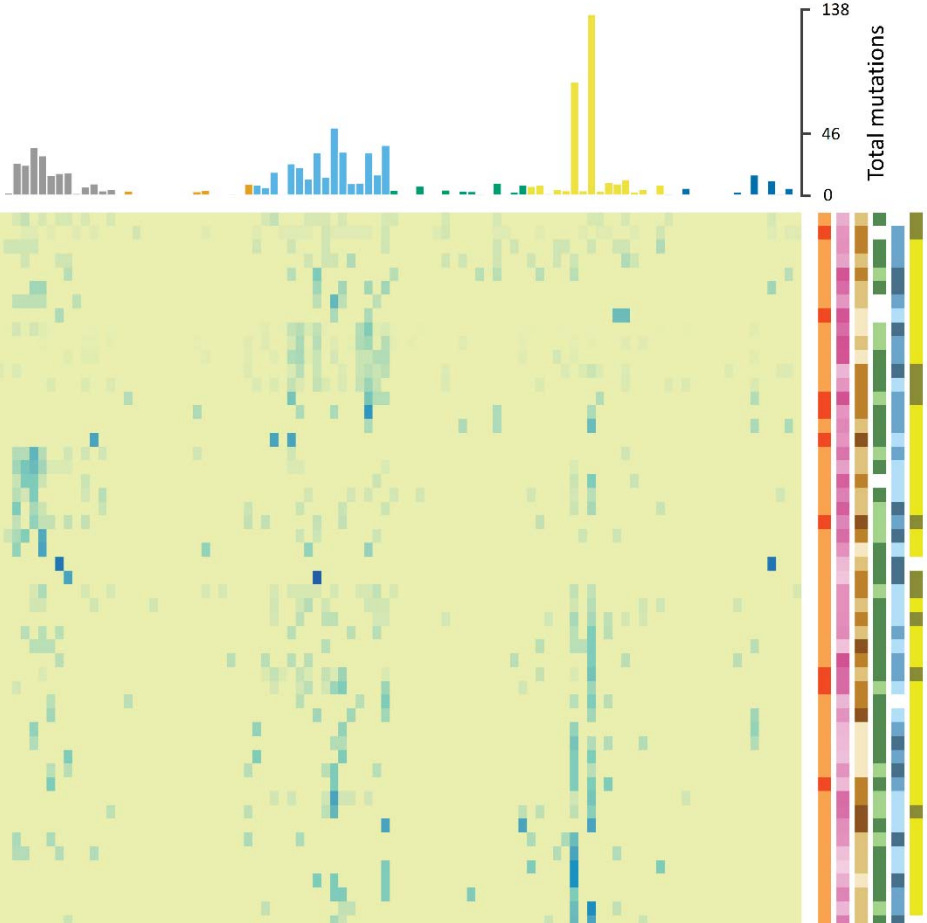

A
*
*
% of R2
0.0
0.2
0.4
Age
Skin photoype
Sun Damage
Sunlight body site
Sunlight history
Sex
MC1R
B
Number of mutations
0
10
50
100
150
250
400
Skin phototype I
Skin phototype II
Skin phototype III
Skin phototype IV
0
10
20
30
40
50
60
70
80
90
100
Age (years)



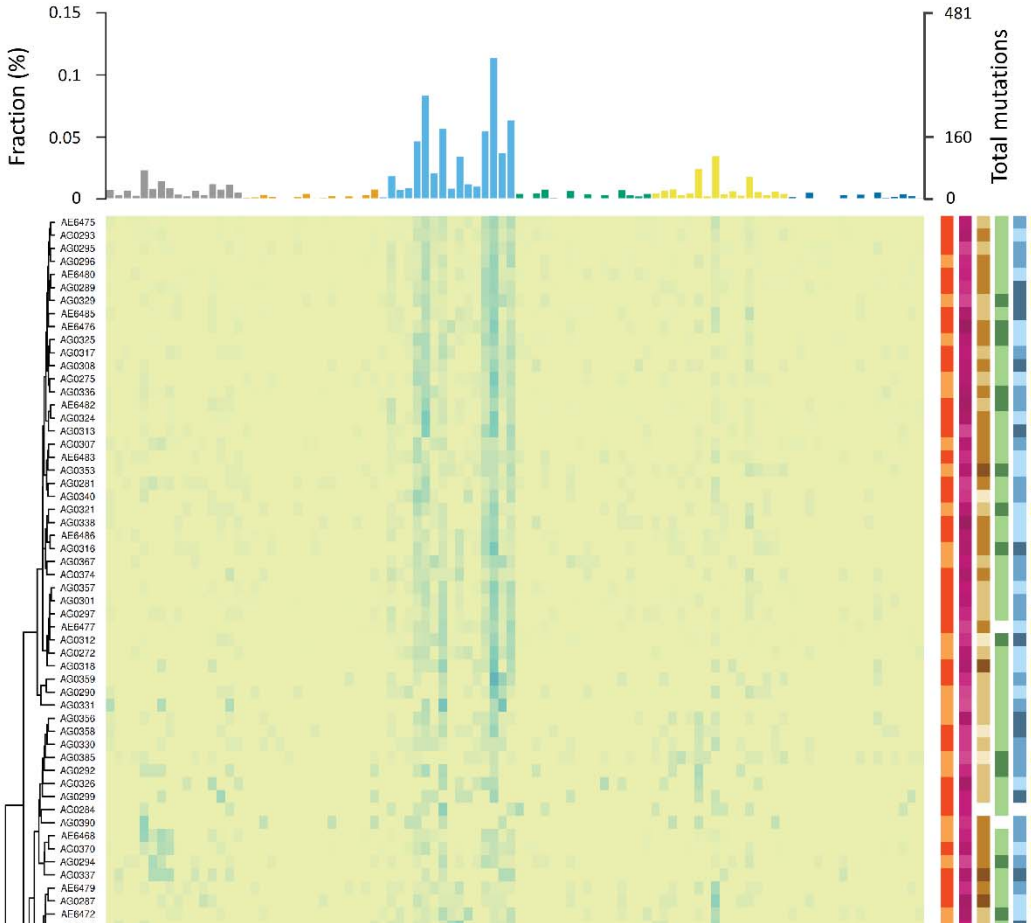


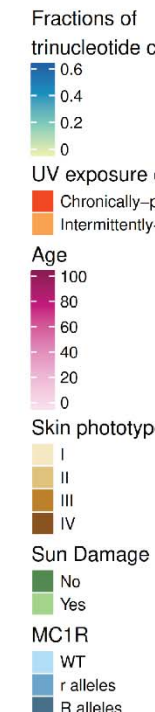


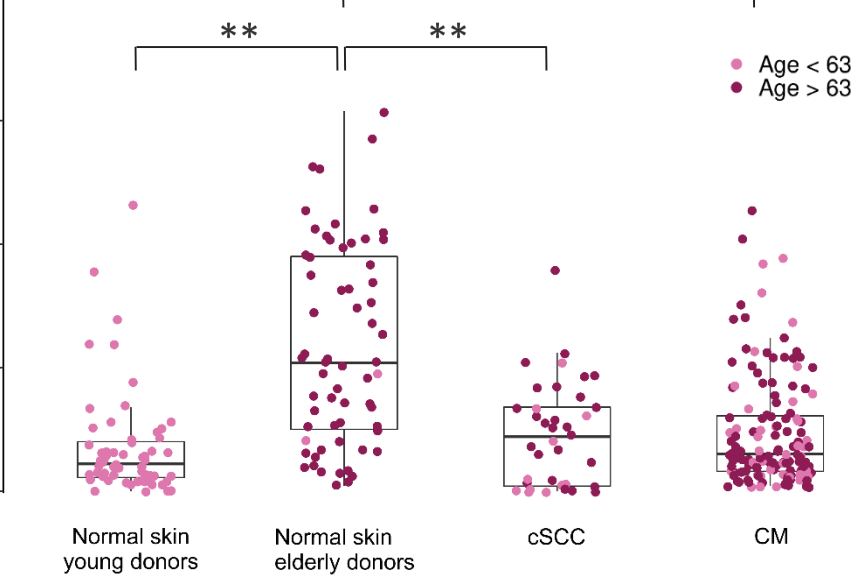
Young donors (Age < 63 years)

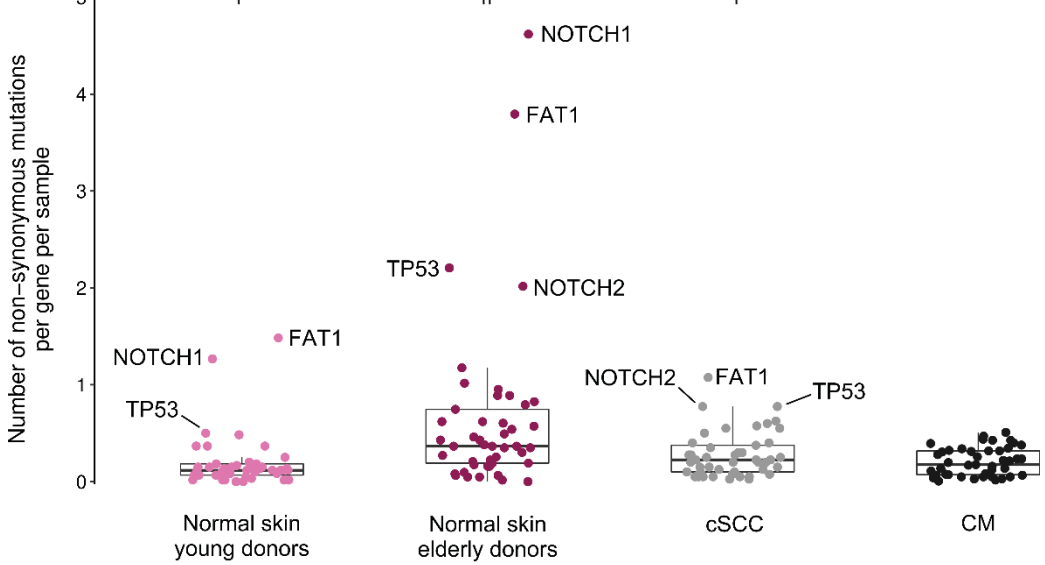


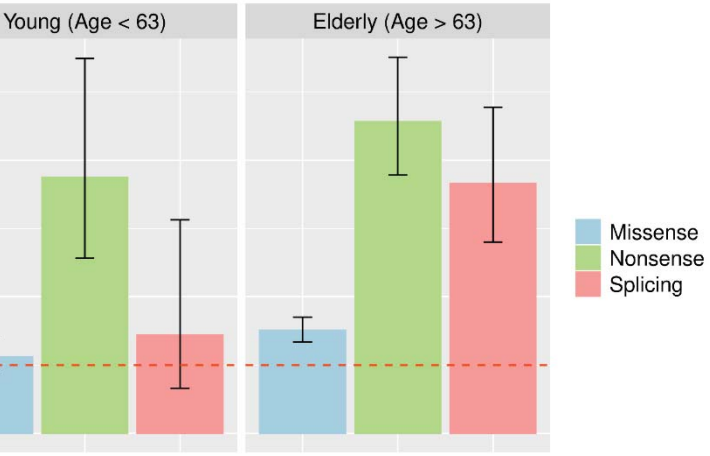
Elderly donors (Age > 63 years)

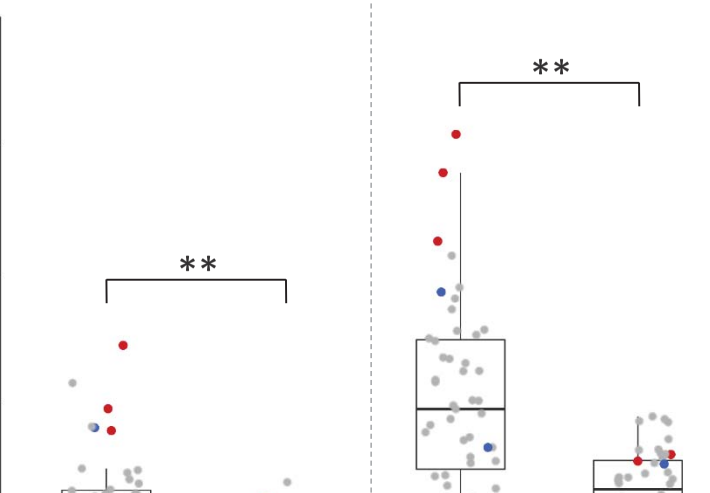





**
**
Age < 63
Age > 63
Normal skin young donors
Normal skin elderly donors
cSCC
CM


** P-valu
Number of non−synonymous mutations per gene per sample
0
1
2
3
4
NOTCH1
FAT1
TP53
NOTCH2
NOTCH1
FAT1
TP53
NOTCH2
FAT1
TP53
Normal skin young donors
Normal skin elderly donors
cSCC
CM


Young (Age < 63)
Elderly (Age > 63)
Missense
Nonsense
Splicing


**
**

D


Missense
Nonsense + Splicing
Indels
IL7R
HRAS
CRNKL1
RB1
TERT
RBM10
PTEN
PTCH1
PREX2
PPP6C
MECOM
KRAS
GRM3
ARID1A
NF1
RAC1
NEBL
EPHA2
PLCB1
NOTCH3
EZH2
KMT2B
PTPRK
NRAS
FGFR3
CDKN2A.p14arf
STAT5B
BAI3
PTPRB
ADAMTS18
SF3B1
SPHKAP
SCN1A
PPP1R3A
FAT2
BRAF
ZNF750
PIK3CA
GRIN2A
SALL1
ADAM29
ARID2
CDKN2A.p16INK4a
TP53
NOTCH2
NOTCH1
FAT1

Supplemental Figures

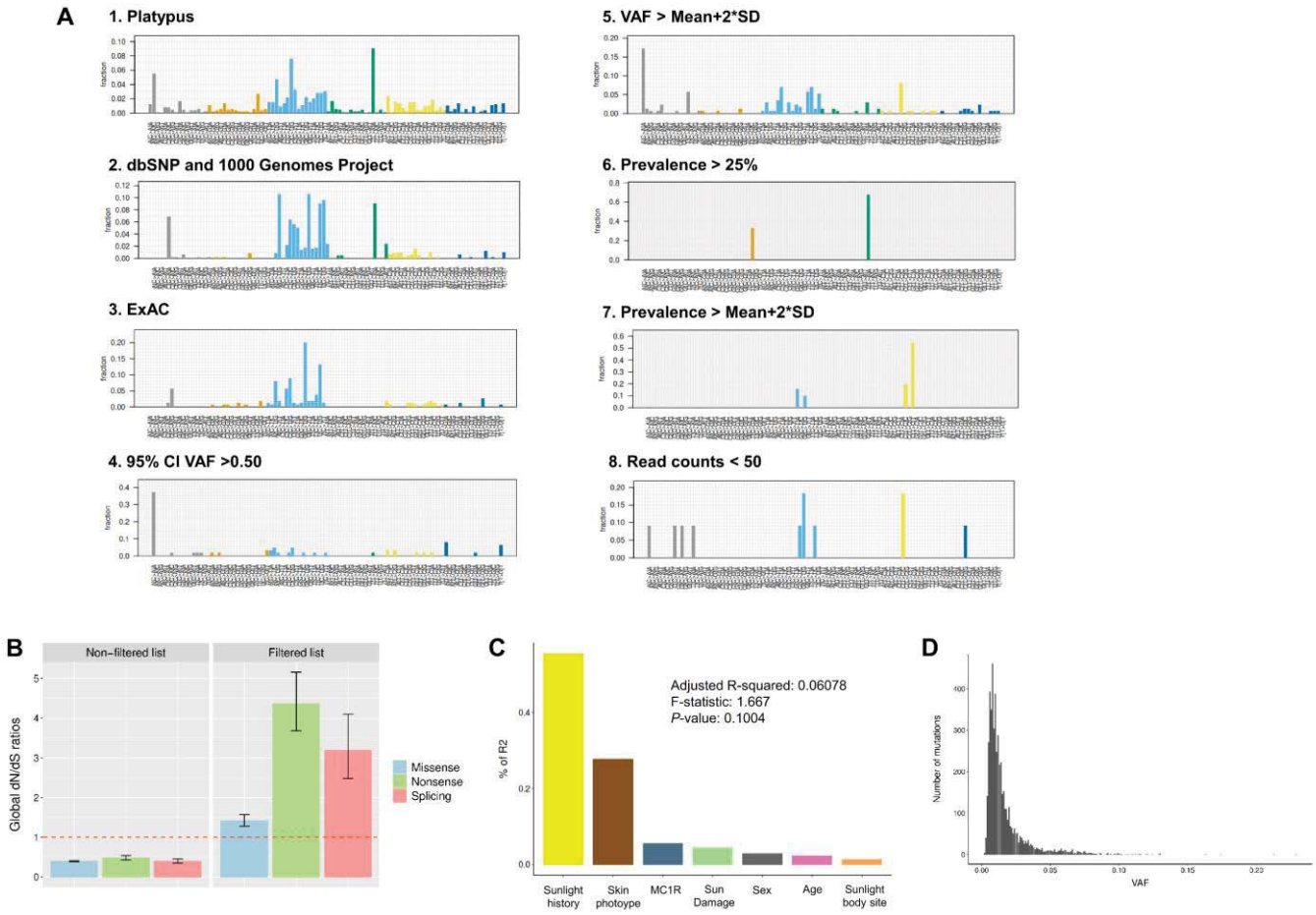


Figure S1. Evaluation of variant calling and filtering, Related to Figures 2 to 5. (A) Spectra of mutation sets removed after applying a specific filtering step. All mutational spectra are very different from the typical UV-related mutational spectrum, indicating that the filtered variants are unlikely to be real somatic mutations. **(B)** Global dN/dS ratios estimated before and after mutation filtering called with Mutect2 tumour-only mode. The global dN/dS << 1 denotes contamination of germline variants and/or technical artefacts in the non-filtered dataset of somatic mutations. This problem seems to be solved after applying the different filtering steps (dN/dS > 1). Error bars denote 95% confidence interval. **(C)** Results of applying a log-linear regression model in the non-filtered mutation dataset for predicting the number of mutations per sample. The low variance explained by the model (adjusted-R^2 = 6.08%) denotes that the non-filtered list of mutations includes a large number of likely false positive calls. **(D)** Histogram of somatic mutations identified by variant allele frequency (VAF). Most somatic mutations remain in a subclonal state with low VAFs (VAF << 5%).

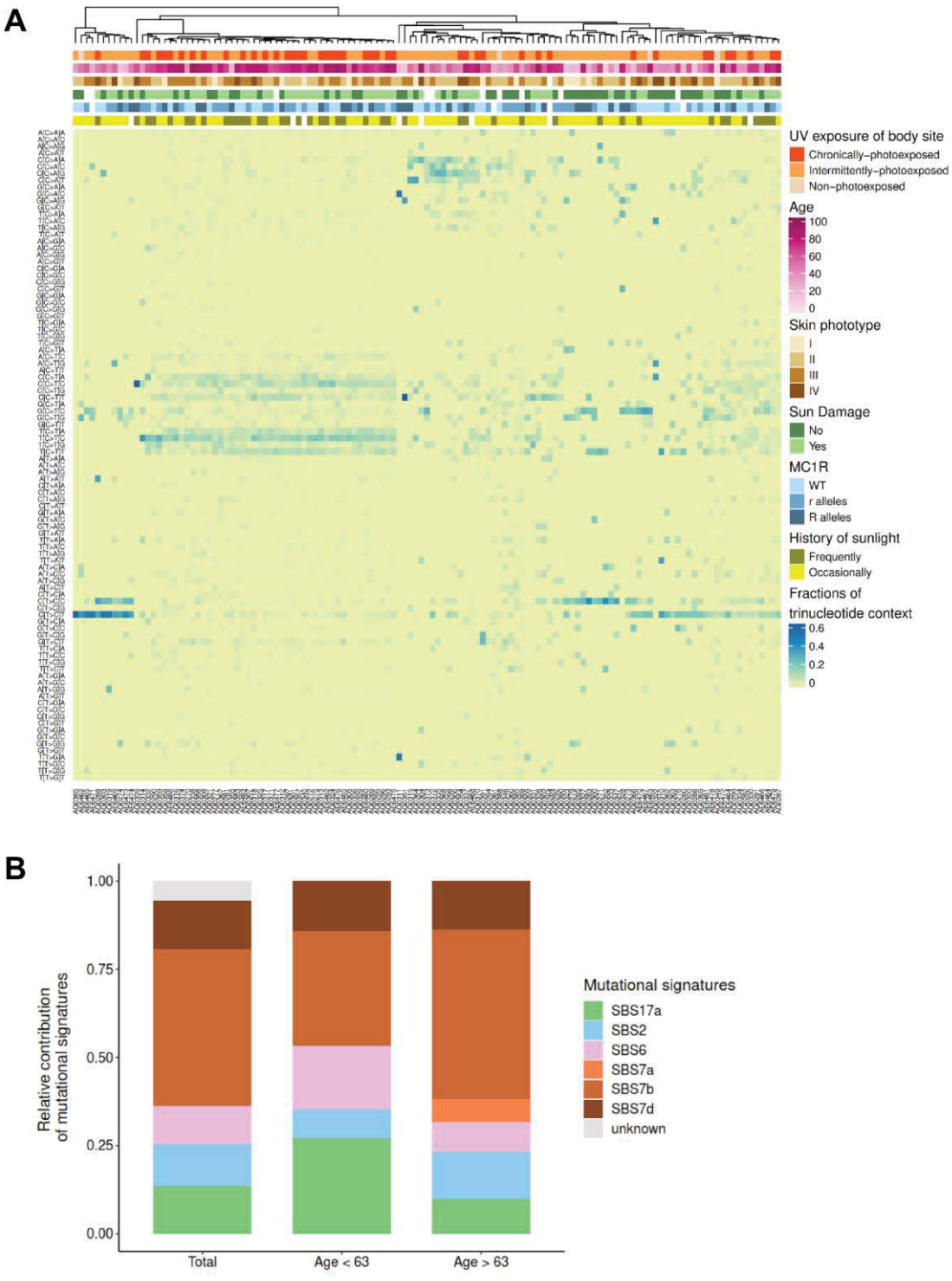


Figure S2. Mutational spectra in normal skin, Related to Figure 4. (A) Heatmap showing the fraction of each 96-mutation type per sample. Clinical and demographic characteristics are presented above each sample. **(B)** Percentage of substitutions attributed to each one of the six mutational signatures for all mutations from all 127 samples together (Total), as well as for all mutations included in each age subgroup.

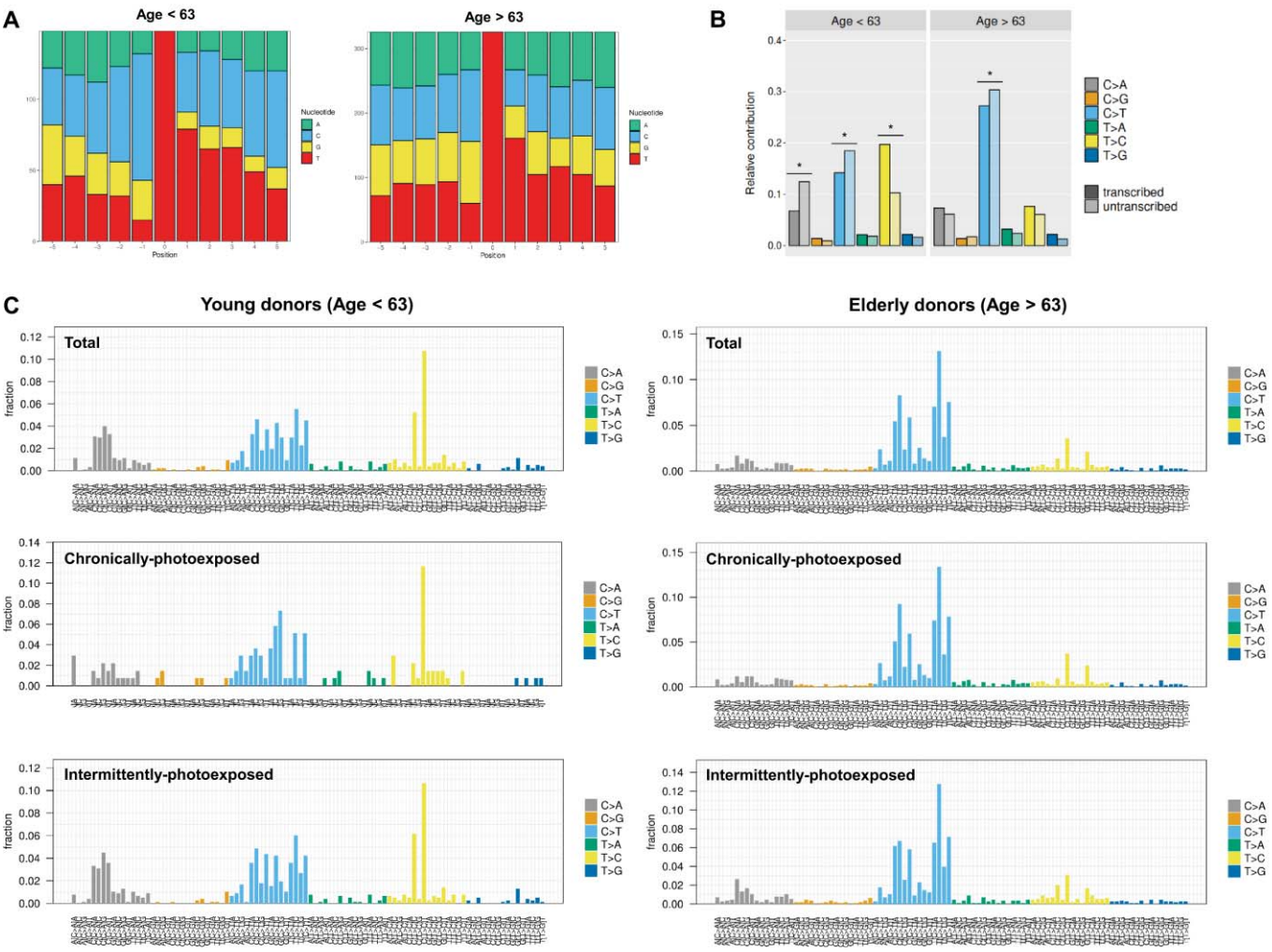


Figure S3. Age-related mutational spectra in normal skin, Related to Figure 4. (A) Local mutational context of T>C substitutions in samples biopsied from young and elderly donors. **(B)** Relative number of each substitution type present on the transcribed (dark shading) and untranscribed strand (light shading) in samples biopsied from young and elderly donors. Asterisks indicate significant transcriptional strand asymmetries (Poisson test). **(C)** 96-barplot depicting the number of mutations observed at each trinucleotide context taking together all samples biopsied from young and elderly individuals (Total), as well as splitting samples of each age group by the body site pattern of sun exposure (Chronically- and Intermittently-photoexposed).

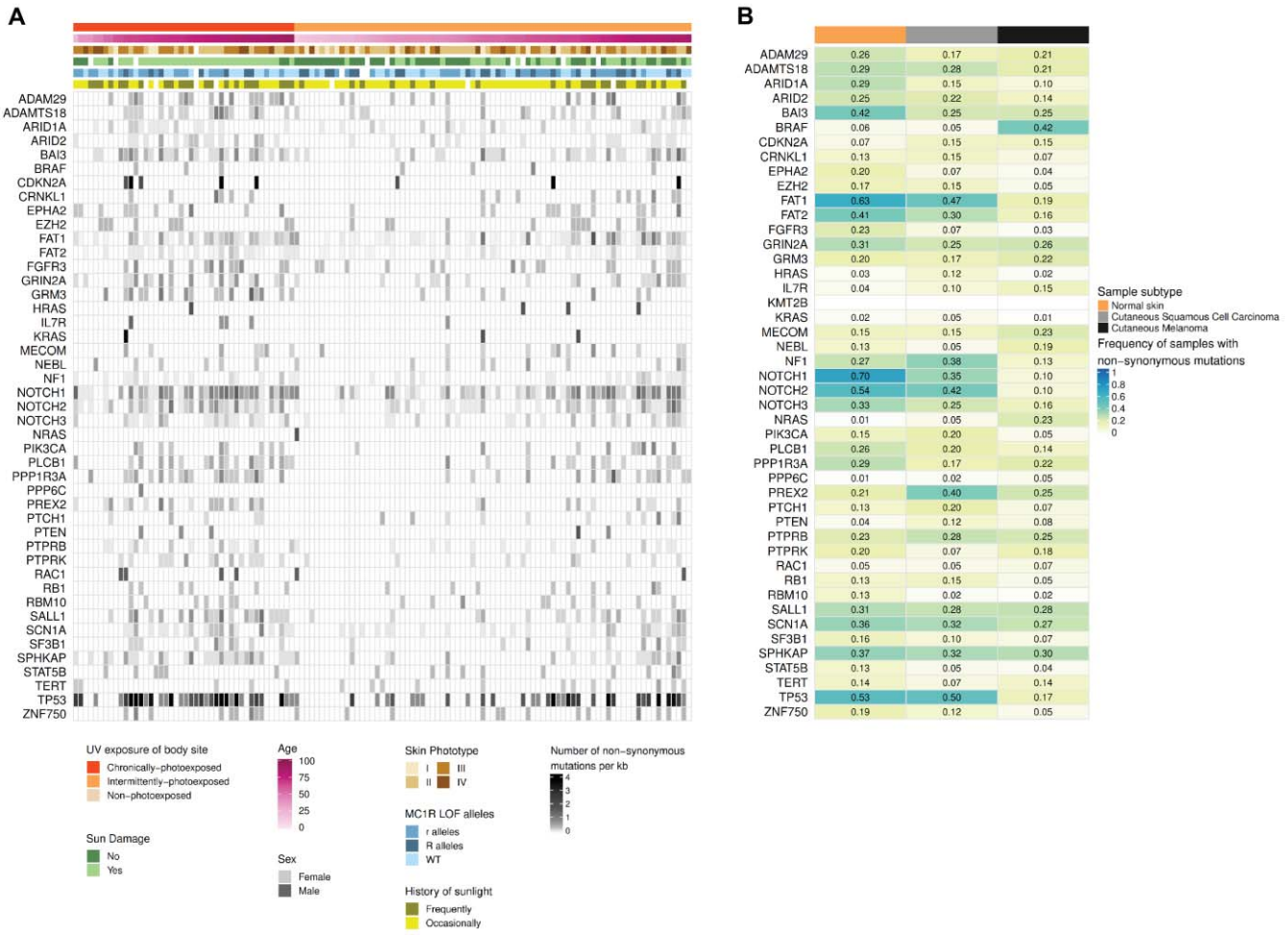


Figure S4. Occurrences of somatic mutations in the 46 cancer genes across samples, Related to Figure 5. (A) Heatmap showing the distribution of recurrent non-synonymous mutations per coding kilobase of sequence for each one of the 46 genes targeted across all normal skin samples. Clinical and demographic characteristics are presented above each sample. The *KMT2B* gene is not included in this plot since no non-synonymous mutation was found across samples. **(B)** Percentage of normal skin samples, as well as in cSCC and melanoma tumours, carrying at least one non-synonymous mutation in each gene.

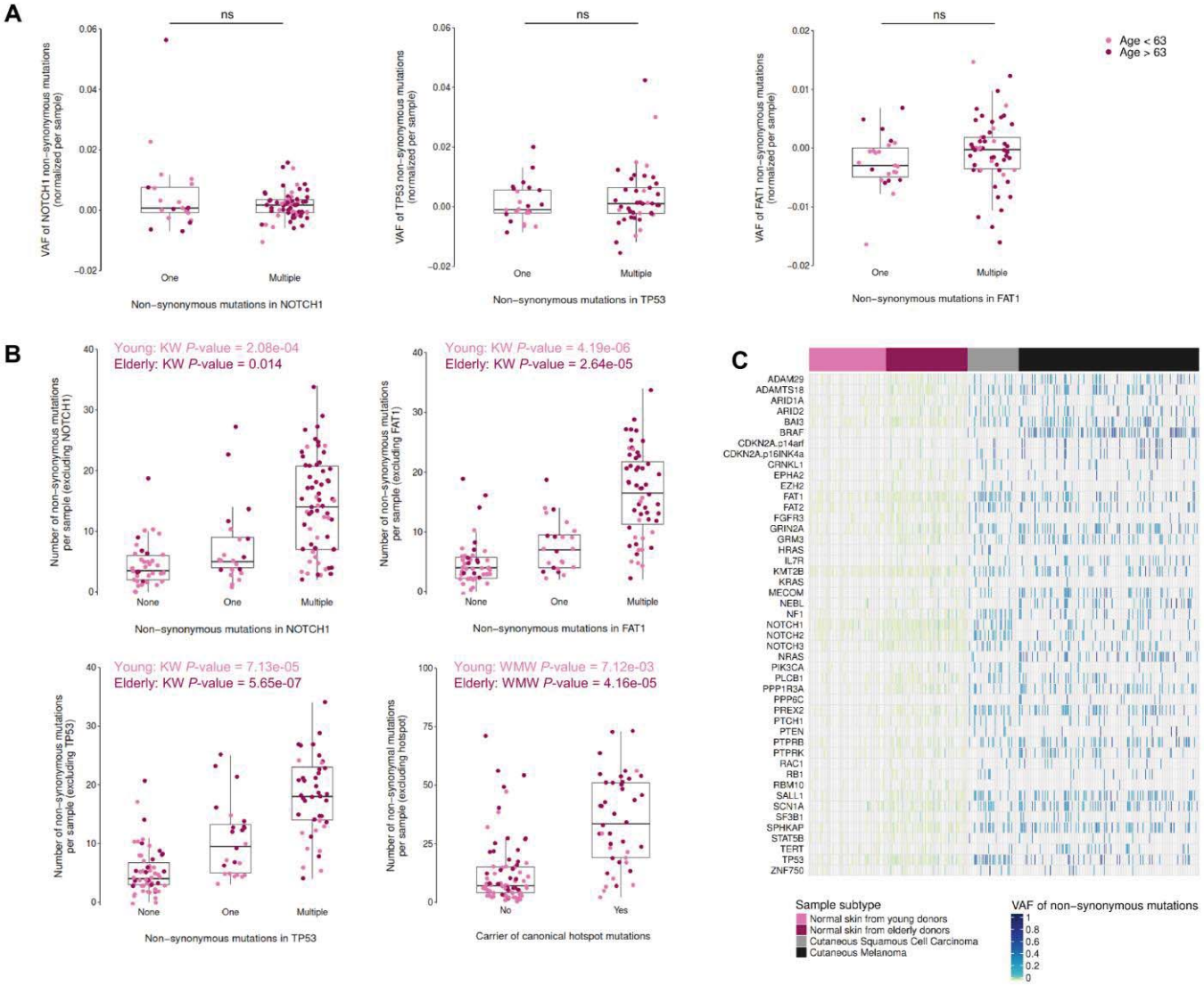


Figure S5. Clonal expansion of clones with oncogenic mutations, Related to Figure 5. (A) Distribution of VAFs for *NOTCH1*, *TP53* and *FAT1* in normal skin samples carrying only one or multiple mutations in the respective gene. For samples with multiple mutations in one of these three genes, the mean VAF for all mutations per gene is plotted. Each dot represents a sample and is coloured according to the donor's age. In each panel, a Wilcoxon-Mann-Whitney test is used for testing differences among groups. **(B)** Number of non-synonymous mutations per sample in normal skin samples non-carriers or carriers of one or multiple non-synonymous mutations in *NOTCH1*, *TP53* and *FAT1*, as well as in normal skin without or with canonical hotspot mutations. Each dot represents a sample and is coloured according to the donor's age. For avoiding the confounding effects of age, samples were stratified according to donor's age for statistical analyses. In panels comparing more than two groups, a Kruskall-Wallis (KW) test is used for testing differences among groups. In panels comparing two groups, a Wilcoxon-Mann-Whitney (WMW) test is used for testing differences among groups. **(C)** Heatmap showing the mean VAF of all non-synonymous mutations found per gene across all normal, cSCC and melanoma samples.

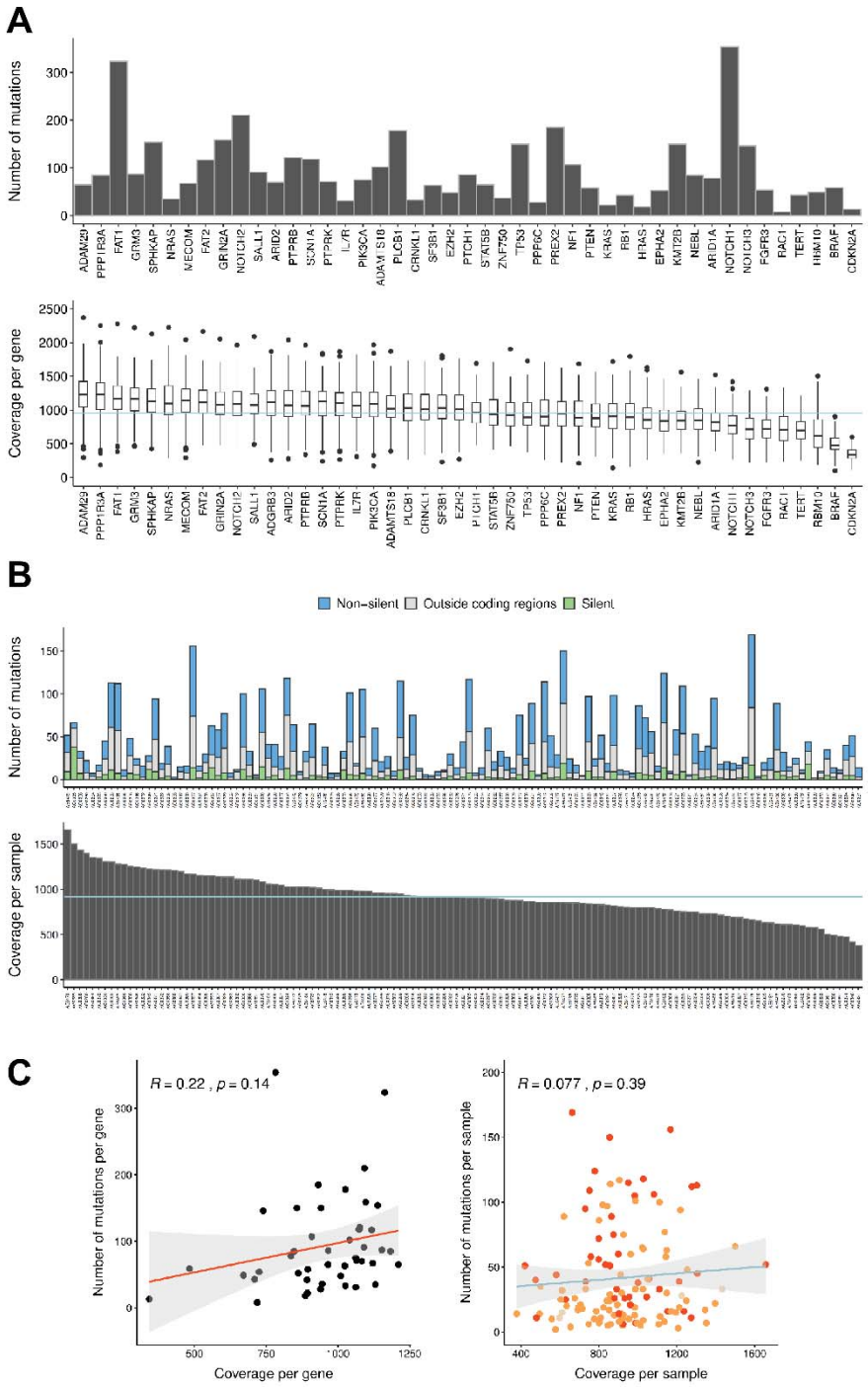


Figure S6. Coverage and mutational burden across genes and samples, Related to Figures 2 to 5. (A) Plot showing the number of mutations per gene across all samples (bar plot, top) and the mean coverage per gene and sample (box plot, bottom). Genes in the x-axis sorted by mean coverage across samples. Blue line indicates the mean coverage across all samples. **(B)** Plot showing the number of mutations per sample (bar plot, top) and the mean coverage per sample (bar plot, bottom). Samples in the x-axis sorted by mean coverage across all sequenced regions. Blue line indicates the mean coverage across all samples. **(C)** Scatter plots showing the coverage and number of mutations per gene (left) and per sample (right). These plots show that coverage did not significantly influence the number of mutations found across genes and/or across samples.

Supplemental Tables

Table S1. Demographic and clinical data of all Spanish donors, Related to Figures 1 to 5.

		Pattern of sunlight exposure of normal skin samples							
		Chronic (N = 44)		Intermittent (N = 79)		Unexposed (N = 4)		Total (N = 127)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age (years)		69.86	15.99	52.16	20.91	45.00	30.32	58.07	21.35
		N	%	N	%	N	%	N	%
Sex	Females	12	27.27	47	59.49	2	50.00	61	48.03
	Males	32	72.73	32	40.51	2	50.00	66	51.97
	Unknown	0	0.00	0	0.00	0	0.00	0	0.00
Fitzpatrick skin type ‡	I	3	6.82	11	13.92	1	25.00	15	11.81
	II	13	29.55	34	43.04	3	75.00	50	39.37
	III	21	47.73	25	31.65	0	0.00	46	36.22
	IV	6	13.64	9	11.39	0	0.00	15	11.81
	Unknown	1	2.27	0	0.00	0	0.00	1	0.79
***MC1R* genotype**	Wild-type	18	40.91	28	35.44	3	75.00	49	38.58
	r carrier	18	40.91	26	32.91	0	0.00	44	34.65
	R carrier	7	15.91	22	27.85	1	25.00	30	23.62
	Unknown	1	2.27	3	3.80	0	0.00	4	3.15
History of sun exposure ¥	Occasional	18	40.91	54	68.35	4	2.17	76	59.84
	Frequent	22	50.00	21	26.58	0	0.00	43	33.86
	Unknown	4	9.09	4	5.06	0	97.83	8	6.30
Sun damage in the skin area	No	6	13.64	46	58.23	3	2.17	55	43.31
	Yes	35	79.55	27	34.18	0	0.00	62	48.82
	Unknown	3	6.82	6	7.59	1	97.83	10	7.87

N, number of individuals; %, percentage of individuals per group among the total

‡ Fitzpatrick's skin type classification is based on pigmentation traits (skin, hair and eye colour) and sun sensitivity-related traits (ability to tan versus tendency to burn, and freckling degree)

¥ History of sun exposure is based on occupancy, outdoor sport activity, and sunbed use

Table S2. Log-linear modelling of the accumulation of somatic mutations in normal skin, Related to Figure 3

Variable	Categories	β	SE	P-value ‡	R^2 (%)
Age		0.028	0.004	**1.46E-09**	55.17
Sex	Female	reference	-	-	3.89
	Male	-0.109	0.165	0.51	
Sunlight exposure of body site	Chronic	reference	-	-	7.83
	Intermittent	-0.171	0.179	0.34	
Sun damage	No	reference	-	-	7.95
	Yes	0.187	0.168	0.27	
Skin phototype	I	reference	-	-	17.93
	II	-0.387	0.266	0.11	
	III	-0.414	0.253	0.15	
	IV	-1.220	0.312	**1.74E-04**	
MC1R genotype	wild-type	reference	-	-	2.00
	r carrier	0.104	0.180	0.56	
	R carrier	-0.226	0.191	0.24	
History of sunlight exposure	Frequent	reference	-	-	5.23
	Occasional	-0.189	0.161	0.24	

β, coefficients; SE, standard error; R^2, percentage of relative contribution of each predictor to the total variance

‡ *P*-value for the multivariate log-linear model